Energy Conversion and Management 99 (2015) 62–71

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

One-year-ahead energy demand estimation from macroeconomic variables using computational intelligence algorithms

S. Salcedo-Sanz [a,*], J. Muñoz-Bulnes [a], J.A. Portilla-Figueras [a], J. Del Ser [b]

[a] Department of Signal Processing and Communications, Universidad de Alcalá, Madrid, Spain
[b] OPTIMA Area, Tecnalia Research & Innovation, Zamudio, Spain

ARTICLE INFO

Article history:
Received 30 July 2014
Accepted 29 March 2015
Available online 24 April 2015

Keywords:
Energy demand estimation
Computational intelligence
Extreme Learning Machines
Harmony Search

ABSTRACT

This paper elaborates on a problem of one-year ahead estimation of energy demand based on macroeconomic variables. To this end, two different Computational Intelligence approaches are herein evaluated: (1) a modified Harmony Search (HS) optimization algorithm with an exponential prediction model and (2) an Extreme Learning Machine (ELM). In the case of the HS, a feature selection of the best set of features for the prediction is carried out jointly with the optimization of the model's parameters. On the other hand, the ELM will be tested with and without the feature selection carried out by the HS approach. We describe several modifications on the proposed HS, which include a hybrid encoding with a binary part for the feature selection, and a real part to tune the parameters of the prediction model. Other adaptations focused on the HS operators are also introduced. The performance of both approaches has been assessed in a real application scenario, corresponding to the total energy demand estimation in Spain, in which we have 14 macroeconomic variables with history values for the last 30 years, including the recent crisis period starting in 2008. The performance of the proposed HS and ELM models incorporating feature selection is shown to provide an accurate one-year-ahead forecast at a higher prediction's accuracy when compared to previous proposals in the literature. Specifically, the HS and ELM approaches are able to improve the results of a previous approach (based on a genetic algorithm), obtaining an improvement over 15% in this problem of energy demand estimation. As a final experimental evaluation of the proposed algorithm, a similar problem of one-year ahead CO_2 emissions estimation from macroeconomic variables is also tackled, and also in this case the HS and ELM are able to obtain significant improvements over a previous approach based on evolutionary computation, over 10% of improvement in this problem.

1. Introduction

In the last decades energy demand has increased sharply at a worldwide scale, pushed up by the globalization phenomenon, a rapid population growth, an aggressive industrialization of developing countries and the high standard of life in developed nations [1]. In this context it is well known that, as the economy grows, the energy demand increases exponentially, what brings along important environmental issues that may compromise the future of next generations. Currently, 80% of the energy demand in the world is covered by non-renewable energy sources such as coal or petroleum, with more dramatic values of this indicator foreseen at developing countries. Another point to be taken into account lies on the fact that industry is the responsible for more than 50% of the energy demand in the world. Consequently, countries with a growing industrial activity happen to be more energy demanding than other with economies based on alternative sectors. Managing medium and long-term energy demand has become a key problem with impact in all countries' economies and nations' development.

Some years ago different studies predicted an increase of the overall energy demand of more than 50% in the next 20 years, in what seemed an unstoppable process [2–6,9]. However, all these forecasts and projections failed after the deep world crisis that started in 2008. The main issue with energy demand estimation problems at a national level is that they depend on macroeconomic variables, which are annually calculated in the majority of cases. Thus, very few data are usually available for constructing the prediction models for energy demand estimation. Furthermore,

* Corresponding author at: Department of Signal Theory and Communications, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain. Tel.: +34 91 885 6731; fax: +34 91 885 6699.
E-mail address: sancho.salcedo@uah.es (S. Salcedo-Sanz).

http://dx.doi.org/10.1016/j.enconman.2015.03.109
0196-8904/

economies of the world 30 years ago were completely different than the current ones, what restricts the variety of historical macroeconomic indicators that can be considered for energy demand estimation.

Having said this, the first approach to tackle this problem was proposed in [3], where a genetic algorithm (GA) was used to obtain the parameters of an exponential prediction model. Specifically, the model proposed in [3] is based on four input macroeconomic variables (Gross Domestic Product or GDP, population, import size and export size) for Turkey, with data recorded from the early 80s to the first years of 2000. The prediction of the energy demand was done for the same year than the input variables (i.e. features' importance is studied for the same year than that of the energy demand, instead of considering the prediction over a given time horizon). Linear and exponential models were considered, whereas the GA was proposed to be a basic binary algorithm, with standard crossover, flip mutation, and a tournament selection. The objective function to be optimized was a measure of mean quadratic error between the real data and the result given by the model, computed over a training set, i.e. a fraction of the available data. With the obtained models it was proven that energy demand in the future could be estimated by projecting variations in the affecting factors (input variables). In this case, future projections predicted a continuous increase of the energy demand in Turkey for the forthcoming 20 years.

The majority of the subsequent literature has since then focused on testing the performance of different evolutionary-type algorithms when applied to this problem, such as Particle Swarm Optimization (PSO) [4,5]) or hybrid approaches based on PSO and Ant Colony Optimization (ACO) [7]. Another hybrid approach blending together PSO and GA has been recently reported in [6,8,10] for energy demand estimation in China. Other approaches have instead elaborated on prediction models from a different approach than the exponential ones used in [3]. Thus, in [11] several new models based on logarithmic and alternative exponential functions are used, optimized by a real-encoding genetic algorithm. All these previous approaches consider a reduced number of affecting factors (input variables or features) from which the obtained projections show a sustained increase of the energy demand in next years. In all cases the training years do not include data beyond 2005, i.e. all years thereafter are missing important events expected to impact on the quality of the performed prediction (e.g. the 2008 global crisis).

In this paper energy demand estimation is tackled from a novel perspective, which combines evolutionary solvers and neural computation algorithms towards an efficient solving methodology. First, we focus on a one-year-ahead energy demand prediction problem: this must be regarded as a major difference with previous approaches where energy estimation is analyzed (i.e. they relate input variables and energy demand, all taken at the same year). In addition, we consider a higher number of predictive (input) variables than previous approaches, with a feature selection procedure to yield the best set of input variables that must be considered for the predictive model. On this purpose we propose to use the Harmony Search (HS) algorithm [12] – a recent evolutionary optimization approach based on mimicking the music generation and improvisation processes – which has obtained very good results in a number of applications [13]. The manuscript describes the proposed approach thoroughly and analyzes its performance when applied with an exponential prediction model to the one-year-ahead energy demand forecast in Spain. The study is further extended by considering a novel neural computation approach – Extreme Learning Machine (ELM) – as the predictive model, which is applied to the complete spectrum of available input variables, as well as to the best set of features obtained by the HS feature selection. An extension of the problem that considers the prediction of CO_2 emissions from macroeconomic variables is also discussed in this work.

The rest of the paper is structured as follows: Section 2 formally describes the problem under consideration, with specific remarks made on the importance of feature selection in prediction problems tackled by means of computational intelligence algorithms. Next, Section 3 describes the fundamentals of the HS algorithm, with details on the used specific encoding and objective function. The main characteristics of the ELM model are also summarized in this section. Section 4 discusses the performance of the proposed algorithms in a real case of energy demand prediction in Spain, for which a comparison with alternative algorithms in the literature is presented. An extension of the problem to a similar case of one-year ahead CO_2 emissions estimation is also discussed in this Section. Finally, Section 5 closes the paper by drawing some ending conclusions.

2. Problem definition

Let us consider a time series $\mathbf{E} \triangleq \{E(t)\}_{t=1}^{n}$ of past energy demands for a given country, with n discrete values corresponding to different years; and a set of m predictive variables $\mathbf{X} = \{X_1(t), \ldots, X_m(t)\}$, with $t = 1, \ldots, n$. A model $\mathcal{M}$ provides an estimation $\widehat{\mathbf{E}}$ for $\mathbf{E}$. The problem tackled in this paper consists of finding the best set of $m' \leqslant m$ features out of the m possible variables in $\mathbf{X}$, as well as the values for the components/parameters of the model $\mathcal{M}$ such that a given objective function – usually related to the similarity of the model output to the real energy demand values – is optimized. In this case, we consider that such a function is given by the mean squared error between the observed values and the predicted ones, which is to be minimized, i.e.

$$f(\mathcal{M}) = \frac{1}{n^*}\sum_{j=1}^{n^*}\left(E(j) - \widehat{E}(j)\right)^2, \qquad (1)$$

where n^* is the size of a reduced training sample ($n^* < n$).

The above formulation corresponds to a class of the so-called Feature Selection (FS) problem. Feature selection is an important task in supervised classification and regression problems because irrelevant features, used as part of the training procedure can unnecessarily increase the cost and running time of a prediction system, as well as degrade its generalization performance [14]. FS problems can be approached under two different schemes: the first attempts at identifying an appropriate set of features independently of the performance of the model $\mathcal{M}$, which preserve most of the information provided by the original data. This approach is known as the *filter* method for feature selection [14], which is outlined in Fig. 1(a) for the sake of completion.

The second FS approach directly selects a subset of m' features out of the total available in such a way that the performance of the model $\mathcal{M}$ is improved or, at least, not degraded. This approach is usually known as *wrapper* method for the FS. These wrapper methods result to be in general more powerful than filter approaches at the usual penalty of an increased computational cost [15,16]. Fig. 1(b) shows an outline of the wrapper method for feature selection. The search of the best subset of input variables can be performed by means of any search algorithm such as hill-climbing and greedy or evolutionary solvers.

3. Materials and methods

As has been advanced in the introduction, this paper evaluates a number of Computational Intelligence approaches when applied to the one-year-ahead energy demand prediction based on