Energy Economics 47 (2015) 37–41

Contents lists available at ScienceDirect

Energy Economics

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

Forecasting short-term electricity consumption using a semantics-based genetic programming framework: The South Italy case

Mauro Castelli [a,*], Leonardo Vanneschi [a], Matteo De Felice [b]

[a] ISEGI, Universidade Nova de Lisboa, 1070-312 Lisboa, Portugal
[b] Energy and Environment Modelling Technical Unit, ENEA, Casaccia R. C., Rome, Italy

ARTICLE INFO

Article history:
Received 13 January 2014
Received in revised form 30 September 2014
Accepted 5 October 2014
Available online 1 November 2014

JEL classification:
C15
C18
C44
C45
C53
C55

Keywords:
Forecasting
Electricity demand
Genetic programming
Semantics

ABSTRACT

Accurate and robust short-term load forecasting plays a significant role in electric power operations. This paper proposes a variant of genetic programming, improved by incorporating semantic awareness in algorithm, to address a short term load forecasting problem. The objective is to automatically generate models that could effectively and reliably predict energy consumption. The presented results, obtained considering a particularly interesting case of the South Italy area, show that the proposed approach outperforms state of the art methods. Hence, the proposed approach reveals appropriate for the problem of forecasting electricity consumption. This study, besides providing an important contribution to the energy load forecasting, confirms the suitability of genetic programming improved with semantic methods in addressing complex real-life applications.

1. Introduction

Load forecasting is the task of predicting electricity demand on different time scales, in minutes (very short-term), hours/days (short-term), and months and years (long-term). This information can be used to plan and schedule operations on power systems (dispatch, unit commitment, network analysis) in a way to control the flow of electricity in an optimal way, with respect to various aspects (like for instance quality of service, reliability, costs). An accurate load forecasting has great benefits for electric utilities and both negative and positive errors lead to increased operating costs. Overestimation of demand leads to an unnecessary energy production or purchase and, on the contrary, underestimation causes unmet demand with a higher probability of failures and costly operations. Several factors influence electricity demand: day of the week and holidays (the so-called "calendar effects"), special or unusual events, economic situation and weather conditions. In warm countries, the last factor is particularly critical during summer, when the use of refrigeration, irrigation and air conditioning is more common than in the rest of the year.

With the recent trend of deregulation of electricity markets, energy demand forecasting has gained even more importance. In the market environment, precise forecasting is the basis of electrical energy trade and spot price establishment for the system to gain the minimum electricity purchasing cost.

All these facts show the importance of having reliable predictive models that can be used for an accurate energy demand forecasting. In this paper, the goal is to propose a new and sophisticated computational method that can be used to automatically generate models for making accurate predictions on the energy demand. This method is based on Artificial Intelligence (AI). The application of an AI technique is aimed at overcoming the limitations of traditional statistics based linear regression methods. Although these techniques and models are reliable, they are generally unable to adapt to unusual events, like for instance sudden changes in the weather conditions and varied holiday activities, which form a highly non-linear relationship with the daily load. Hence, their load predictions in the presence of such events are often not as satisfactory as desired, and consequently, more sophisticated methods are needed in order to accurately map the correlation between all the variables. AI methods for forecasting have shown an ability to give better performance in dealing with non-linearity and other difficulties in modeling of time series. Their advantage lies mainly in the fact that they do not

* Corresponding author. Tel.: +351 213828628; fax: +351 213872140.
E-mail addresses: mcastelli@isegi.unl.pt (M. Castelli), lvanneschi@isegi.unl.pt (L. Vanneschi), matteo.defelice@enea.it (M. De Felice).

http://dx.doi.org/10.1016/j.eneco.2014.10.009

require any complex mathematical formulations or knowledge of predetermined relationships between inputs and outputs.

In this study, the focus is on Genetic Programming (GP) (Koza, 1992; Vanneschi and Poli, 2012), that is one of the youngest AI techniques. In particular, a recently defined and very promising variant of standard GP is proposed, integrating the concept of semantic awareness in the algorithm used to generate predictive models.

To analyze the appropriateness of the proposed computational method for energy Short-Term Load Forecasting (STLF), the energy consumption in a particular area of Italy has been investigated. As reported by the main Italian energy provider, electricity consumption in Italy is rising. Early data on the trend of the demand in 2010 showed a 1.8% increase compared to 2009, the highest positive variation from 2007 to date. The total energy required in Italy equaled 326.2 billion kWh. In 2010, the electricity demand was met at 86.5% with the national production and for the remaining part (13.5%) with the balance of electricity exchanged with foreign countries. This amount of imported energy makes Italy the world's biggest importer of electricity. Due to its reliance on expensive fossil fuels and imports, Italians pay approximately 45% more than the EU average for electricity.

The area under exam includes the regions of southern Italy. South Italy regions have been taken into account because they represent a very interesting test case, given the important contribution that they have in building the global Italian energy demand and considering the high variability of energy demand in this area in recent years. Moreover, this area is challenging for the STLF problem: while AI-based models and traditional statistics-based models are able to produce accurate predictions for a large part of the Italian area, the regional area considered in this study presents particular features that often cause the forecasting models to not produce satisfactory predictions. The main motivations for the high variability in terms of energy demand in this area are discussed in Section 3.1.

The paper is organized as follows: Section 2 presents the variant of GP proposed in this study for addressing the STLF problem. Section 3 describes the data that have been considered in this study and reports experimental results comparing the proposed approach to the standard GP algorithm and other state of the art techniques. Section 4 concludes the paper, highlighting the main contributions of this work. In the final part of the manuscript, appendices contain general introductions of basic concepts for non-experts; more in particular, Appendix A describes the energy load forecasting problem, while Appendix B introduces the standard GP method.

2. Methodology

Models lie in the core of any technology in any industry, be it finance, manufacturing, services, mining, or information technology. The task of data-driven modeling lies in using a limited number of observations of system variables for inferring relationships among these variables. The design of reliable learning machines for data-driven modeling tasks is of strategic importance, as there are many systems that cannot be accurately modeled by classical mathematical or statistical techniques. Reliable learning in the field of Machine Learning (ML) revolves around the notion of generalization, which is the ability of a learned model to correctly explain data that are drawn from the same distribution as training data, but have not been presented during the training process.

Genetic programming (GP) (Koza, 1992; Poli et al., 2008) is one of the youngest paradigms inside the computational intelligence research area called Evolutionary Computation (EC) and consists in the automated learning of computer programs by means of a process mimicking Darwinian evolution. GP tackles learning problems by searching a computer program space for the program that better respects some given functional specifications. In GP a population of computer programs is evolved. That is, generation by generation, GP stochastically transforms populations of programs into new, hopefully better, populations of programs. This process is generally driven by a selection algorithm, mimicking Darwinian natural selection, and by transformation, or genetic, operators, usually crossover and mutation, mimicking the homonymous biological processes (an introduction to GP can be found in Appendix B).

In the last few years, GP has produced a wide set of extremely interesting applicative results, some of which have been defined human-competitive (Koza, 2010). While these results have demonstrated the suitability of GP in tackling real-life problems, research has recently focused on developing new variants of GP in order to further improve its performance. In particular, efforts have been dedicated to an aspect that was only marginally considered up to some years ago: the definition of methods able to exploit semantic awareness of the solutions (Beadle and Johnson, 2009; Jackson and Promoting phenotypic diversity in genetic programming, in: PPSN,, 2010; Krawiec and Lichocki, 2009; Vanneschi et al., 2014). Although there is no universally accepted definition of semantics in GP, this term often refers to the behavior of a program, once it is executed on a set of data. For this reason, in many references, including here, the term semantics is intended as the vector of outputs a program produces on the training data (Moraglio et al., 2012). Although semantics determines what a program actually does, the traditional GP operators, like crossover and mutation described so far ignore this knowledge and manipulate programs only at a syntactic level. Abstraction from semantics allows them to rely on simple, generic search operators, but the main consequence of this choice is that it is difficult to predict how modifications of programs will affect their semantics. Recently, new genetic operators, called geometric semantic genetic operators have been proposed in (Moraglio et al., 2012). These operators, that manipulate programs considering directly their semantic information, have a number of theoretical advantages, compared to the ones of standard GP, the most important one being the fact that they induce a unimodal fitness landscape (Stadler and Towards a theory of landscapes, 1995) on any problem consisting in finding the match between a set of input data and a set of expected target ones. According to the theory of fitness landscapes (Vanneschi, 2004) this should relevantly improve GP evolvability (i.e. the ability of genetic operators to produce offspring that are fitter than their parents) on all these problems. In this section, we report the definition of geometric semantic operators for real functions domains, since these are the operators we will use here. For applications that consider other kinds of data, the reader is referred to Moraglio et al. (2012).

Definition. Geometric semantic crossover

Given two parent functions $T_1, T_2 : \mathbb{R}^n \rightarrow \mathbb{R}$, the geometric semantic crossover returns the real function $T_{XO} = (T_1 \cdot T_R) + ((1 - T_R) \cdot T_2)$, where T_R is a random real function whose output values range in the interval [0, 1].

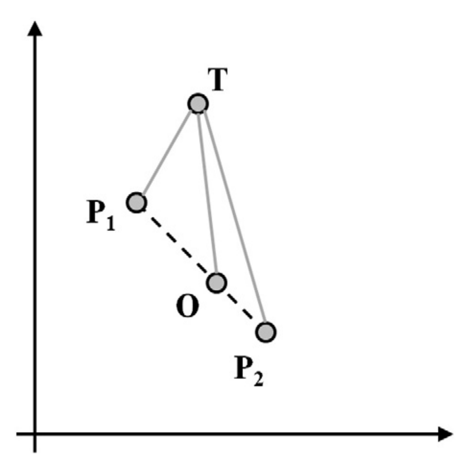

Fig. 1. Graphical representation of a toy bi-dimensional semantic space that we use to give a visual intuition of the fact that geometric semantic crossover produces an offspring that is at least not worse than the worst of its parents. In this simple case, offspring O (which stands between parents P_1 and P_2 by construction) is clearly closer to target T than parent P_2.