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
0140-9883/

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 $T_1, T_2 : \mathbb{R}^n \to \mathbb{R}$, 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].

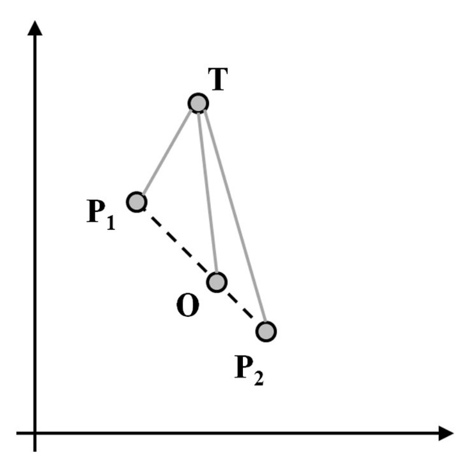

Fig. 1. Graphical representation of a toy bi-dimensional semantic space that we use to give a visual intuition of the fact that geometric semantic crossover produces an offspring that is at least not worse than the worst of its parents. In this simple case, offspring O (which stands between parents P_1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

This operator corresponds to a geometric crossover on the semantic space, in the sense that it produces an offspring that stands between its parents in this space. Even without a formal proof, we can have an intuition of it by considering that the (unique) offspring generated by this crossover has a semantic vector that is a linear combination of the semantics of the parents with random coefficients included in [0, 1] and whose sum is equal to 1. An interesting consequence of this fact is that the fitness of the offspring cannot be worse than the fitness of the worst of its parents. Also in this case we omit the formal proof, but we limit ourselves to give a visual intuition of this property: in Fig. 1 we report a simple two-dimensional semantic space in which we draw a target function T, two parents P_1 and P_2 and one of their, descendants O (which by construction stands in the segment that joins its parents).

Given that in supervised learning applications fitness often corresponds to distance to the target, in Fig. 1, it is immediately clear that O is closer to T, and thus better, than P_2 (which is the worst parent in this case). To constrain T_R in producing values in [0, 1] we use the sigmoid function: $T_R = \frac{1}{1+e^{-T_{rand}}}$, as in (Vanneschi et al., 2013), where T_{rand} is a random tree with no constraints on the output values.

Definition. Geometric semantic mutation

Given a parent function $T:\mathbb{R}^n \rightarrow \mathbb{R}$, the geometric semantic mutation with mutation step ms returns the real function $T_M = T + ms \cdot (T_{R1} - T_{R2})$, where T_{R1} and T_{R2} are random real functions whose output values range in [0, 1].

To constrain T_{R1} and T_{R2} to produce values in [0, 1], we use the sigmoid function exactly as for T_R in the definition of geometric semantic crossover. Reference (Moraglio et al., 2012) formally proves that this mutation operator induces a unimodal fitness landscape for all the problems consisting in mapping sets of input data into targets, like for instance classification or regression. However, even though without a formal proof, it is not difficult to have an intuition of it, considering that each element of the semantic vector of the offspring is a “weak” perturbation of the corresponding element in the parent's semantics. We informally define this perturbation as “weak” because it is given by a random expression centered in zero (the difference between two random trees with outputs in [0, 1]). Nevertheless, by changing parameter ms, we are able to tune the “step” of the mutation and thus the importance of this perturbation.

It is important to point out that at every step of one of these operators, an offspring contains the complete structure of the parents, plus one or more random trees as its subtrees and some arithmetic operators: the size of each offspring is thus clearly much larger than the one of their parents. The work we proposed in Vanneschi et al. (2013) explained how the rapid growth of the individuals can be addressed, with a very simple and effective implementation of the GP algorithm. This is the implementation used in this paper.

3. Experimental study

This section begins with a brief description of the main characteristics of the regions of the south of Italy, motivating the fact that the study of these regions is particularly interesting for energy STLF. Subsequently, we describe the data, the used experimental settings and the obtained results for the energy STLF in that geographic area.

3.1. South Italy area: main characteristics

The regions of Italy are the first-level administrative divisions of the state according to NUTS (nomenclature of territorial units for statistics). In this study we have taken into account an area of South Italy that includes the following regions: Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise and Sicily. The reasons for the choice of this particular geographic area reside in certain climatic and economic factors that make it particularly interesting in the context of STLF problem. As mentioned before, the energy demand is linked to various factors and, in particular, to the economic and climatic conditions. In recent years, the geographical area under exam has been characterized by profound transformations that are still changing the economic scenario. Basically, South Italy has moved from an economy that is mainly based on agriculture or otherwise characterized by a shortage of industries to an economy more closely linked to industry. This transformation is obviously reflected in a change in the energy policy. In particular, the forecast models should serve this change, and this variability affects the quality of the forecasting models.

Another factor that makes South Italy an interesting test case is related to climate. This geographical area is characterized by a large temperature difference between the values recorded in the winter and the ones recorded in the summer. Moreover, in recent years, there have been significant climate changes that have increased the variability of the values recorded in the same season (or even in the same month) but in different years. Hence, to obtain reliable forecasting models, it is necessary to consider for this area, both the intra-season and the inter-season temperature variability. This consideration is particularly important for mid-term and long-term forecasting, but it should also be considered in the definition of short-term forecasting models. In fact, in the regional area under investigation, it is not difficult to register differences in terms of temperatures between two consecutive weeks (or days), that can even reach a value of 10 °C. For a complete analysis of the climate-related characteristics and changes of the South Italy area considered in this study the reader is referred to Brunetti et al. (2001).

3.2. Data description

Historical load data and weather information for the South Italy area are collected by TERNA in the years between 1999 and 2010. TERNA S.p.A. (*Rete Elettrica Nazionale*) is an Italian electricity transmission system operator based in Rome, Italy. With 63,500 km of power lines or around 98% of the Italian high-voltage power transmission grid, TERNA is the first independent electricity transmission grid operator in Europe and the sixth in the world based on the size of its electrical grid. TERNA is the owner of the Italian transmission grid and the responsible for energy transmission and dispatching. The aim of the forecasting task studied in this paper is to predict the load at time t providing information until day $t-1$ (one-day ahead forecasting) using the past samples of the load and weather information. Data include temperatures, pressure values, wind speed and other weather related information. Data are collected by using detection devices arranged in the area of interest. Data from 1999 to 2006 have been used during the training phase, while the remaining available data have been used to validate the model on unseen data, hence to assess the quality of the forecasting.

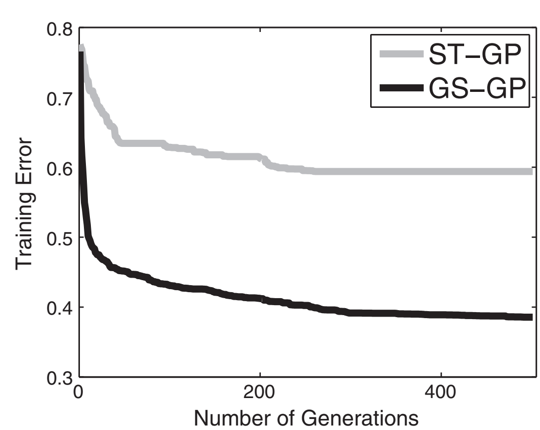

Fig. 2. Median of the training fitness for ST-GP and GS-GP at each generation calculated over 50 independent runs.

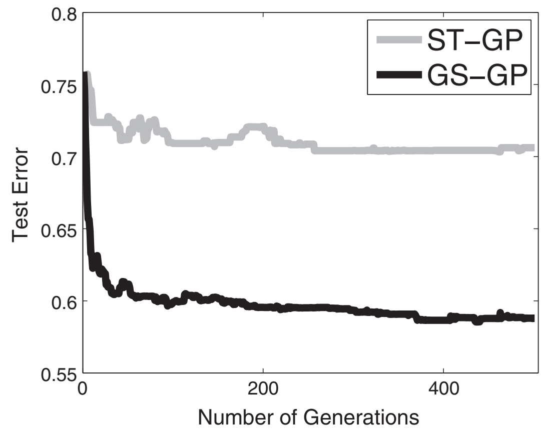

Fig. 3. Median of the test fitness for ST-GP and GS-GP at each generation calculated over 50 independent runs.

3.3. Experimental settings

We tested the proposed implementation of GP with geometric semantic operators (GS-GP from now on) against a standard GP system (ST-GP). A total of 50 runs were performed with each technique, and this is a fundamental aspect given the stochastic nature of the considered systems. All the runs used populations of 200 individuals allowed to evolve for 500 generations. Tree initialization was performed with the Ramped Half-and-Half method (Koza, 1992) with a maximum initial depth equal to 6. The function set contained the arithmetic operators, including division protected as in Koza (1992). Fitness was calculated using the Canberra distance (Lance and Williams, 1967) between predicted and target values. The use of this distance produces fitness values between 0 and 1, where 0 is the optimal fitness (thus the lower the fitness, the better the individual). The terminal set contained 45 variables, each one corresponding to a different feature in the dataset. To create new individuals, ST-GP used standard (subtree swapping) crossover (Koza, 1992) and (subtree) mutation (Koza, 1992) with probabilities equal to 0.9 and 0.1, respectively. For GS-GP, crossover rate was equal to 0.7, while mutation rate was 0.3. The motivation for these different values for the two GP systems is that a preliminary experimental study has been performed (independently for the two systems) for finding the parameter setting able to return the best results. Only the parameter settings that returned the best results for the two systems are presented here. Survival from one generation to the other was always guaranteed to the best individual of the population (elitism).

In the next section, the obtained experimental results are reported using curves of the Canberra distance on the training and test set. In particular, at each generation the best individual in the population (i.e. the one that has the smaller training error) has been chosen and the value of its error on the training and test sets has been stored. The reported curves finally contain the median of all these values collected at each generation. The median was preferred over the mean in the reported plots because of its higher robustness to outliers. The Canberra distance on the training and test set will be in some cases informally indicated as training and test fitness, or training and test error, in the next section for simplicity.

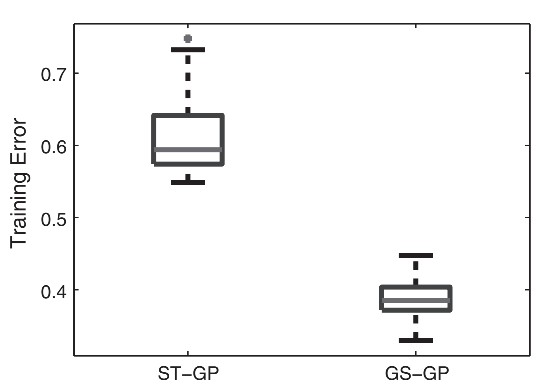

Fig. 4. Training fitness of the best individual produced in each of the 50 runs at the last performed generation.

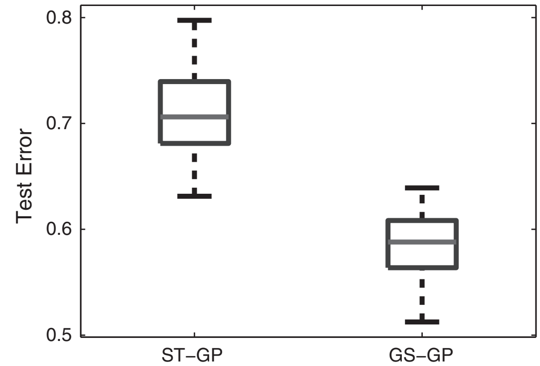

Fig. 5. Test fitness of the best individual produced in each of the 50 runs at the last performed generation.

The results discussed in the next section have been obtained using the GS-GP implementation freely available at http://gsgp.sourceforge.net and documented in Castelli et al. (2014).

3.4. Experimental results: GS-GP vs ST-GP

Figs. 2 and 3 report training and test error for ST-GP and GS-GP against generations.

These figures clearly show that GS-GP outperforms ST-GP on both training and test sets. Figs. 4 and 5 report a statistical study of the training and test fitness of the best individual, both for GS-GP and ST-GP, for each of the 50 performed runs. Denoting by IQR the interquartile range, the ends of the whiskers represent the lowest datum still within 1.5 IQR of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. As it is possible to see, GS-GP produces solutions that, besides being more accurate, also have a lower dispersion with respect to the ones produced by ST-GP. To analyze the statistical significance of these results, a set of tests has been performed on the median errors. As a first step, the Kolmogorov–Smirnov test has shown that the data are not normally distributed and hence a rank-based statistic has been used. Then, the Wilcoxon rank-sum test for pairwise data comparison has been used under the alternative hypothesis that the samples do not have equal medians. The *p*-values obtained are $3.69 \cdot e^{-11}$ when the test fitness of ST-GP is compared to the test fitness of GS-GP and $3.02 \cdot e^{-11}$ when the training fitness of ST-GP is compared to the training fitness of GS-GP. Therefore, when employing the usual significance level $\alpha = 0.01$, we can clearly state that GS-GP produces solutions

Table 1
Experimental comparison between different non-evolutionary techniques and GS-GP. For non deterministic techniques we reported the median of the training error and test error calculated over 50 independent runs.

Method	Training error	Test error
Linear regression Weisberg (2005)	0.67	0.75
Least square regression Seber and Wild (2003)	0.61	0.78
Radial basis function network Haykin (1999)	0.47	0.66
Isotonic regression Hoffmann (2009)	0.64	0.76
SVM polynomial kernel (degree 1) Schölkopf and Smola (2002)	0.53	0.68
SVM polynomial kernel (degree 2)	0.49	0.71
GS-GP	0.39	0.59

Table 2
p-Values given by the statistical test for the experimental comparison between GS-GP and the other studied non-evolutionary techniques.

		LIN	SQ	RBF	ISO	SVM-1	SVM-2
GS-GP	Train	$1.21 \cdot e^{-012}$	$1.21 \cdot e^{-012}$	$2.28 \cdot e^{-9}$	$1.21 \cdot e^{-012}$	$1.21 \cdot e^{-012}$	$1.21 \cdot e^{-012}$
	Test	$1.21 \cdot e^{-012}$	$1.21 \cdot e^{-012}$	$1.58 \cdot e^{-9}$	$1.21 \cdot e^{-012}$	$1.21 \cdot e^{-012}$	$1.21 \cdot e^{-012}$

that are significantly better (i.e. with lower error) than ST-GP both on training and test data.

3.5. Experimental results: GS-GP vs other state-of-the-art techniques

Besides comparing GS-GP with ST-GP, we are also interested in comparing GS-GP with other well-known state-of-the-art methods, to have an idea of the competitiveness of the obtained results. Table 1 reports the values of the training and test errors of the solutions obtained by all the studied techniques including, in the last line of the table, GS-GP.

From these results, it is possible to see that GS-GP performs better than all the other studied methods, both on training as well as on unseen test data.

To assess the statistical significance of these results, the same set of tests described in the previous section has been performed. In this case, a Bonferroni correction for the value of α has been considered, given that the number of compared techniques is larger than two. All the obtained *p*-values relative to the comparison between GS-GP and the other methods are reported in Table 2.

In the table LIN stands for linear regression, SQ stands for least square regression, RBF stands for radial basis function network, ISO stands for isotonic regression, SVM-1 refers to the support vector machines with polynomial kernel of first degree and SVM-2 refers to the support vector machines with polynomial kernel of second degree. According to the results reported in the table, the differences in terms of training and test fitness between GS-GP and all the other considered techniques are statistically significant. These results are a clear indication of the appropriateness of GS-GP as a method to generate predictive models for the STLF problem, at least for the studied test case.

4. Conclusions

Electricity short term load forecasting is important for the power industry, especially in the context of the ongoing deregulation of the electricity market. Proper demand forecasts help the market participants to maximize their profits and/or reduce their possible losses by preparing an appropriate bidding strategy. In this study, the Short-Term Load Forecasting (STLF) problem has been considered, focusing on the energy demand of the South Italy regions. This area has been chosen due to the high variability in its energy demand pattern and for the differences (i.e., weather and economy) that presents with respect to the rest of Italy. Traditional statistics based linear regression methods often need modification to capture the non-linearity in demand signals under the market conditions. To overcome this limitation, systems based on Artificial Intelligence (AI) are becoming more and more common decision making tools for STLF. Among the different AI techniques, Genetic Programming (GP) looks particularly promising, given that it has shown its suitability in dealing with time series analysis so far. In this work, we have proposed a new GP framework that, differently from the standard GP algorithm, integrates semantic awareness in the definition of the genetic operators. To validate the proposed method, called Geometric–Semantic GP (GS-GP), an extensive experimental analysis has been performed, considering the data that cover the period 1990–2010. The reported results have shown that GS-GP is able to produce results that are statistically better than the ones produced by the canonical GP algorithm and by other state-of-the-art techniques. In particular, GS-GP is able to reduce the forecasting error with respect to all the considering techniques, thus generating more accurate and reliable predictive models. This is clear indication of the appropriateness of GS-GP for the STLF problem, at least for the South Italy area test case studied here.

These results are extremely encouraging, and pave the way to a deeper study of GS-GP for STLF. In particular, in the future, we plan to extend our study to other geographic areas, including much larger ones, in which the problem of forecasting energy consumption is particularly delicate or relevant.

Acknowledgments

The authors acknowledge project MassGP (PTDC/EEI-CTP/2975/2012), FCT, Portugal.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.eneco.2014.10.009.

References

Beadle, L., Johnson, C.G., 2009. Semantically driven mutation in genetic programming. In: Tyrrell, A. (Ed.), 2009 IEEE Congress on Evolutionary Computation, IEEE Computational Intelligence Society. IEEE Press, Trondheim, Norway, pp. 1336–1342.
Brunetti, M., Buffoni, L., Mangianti, F., Maugeri, M., Nanni, T., 2001. Climate variations in Italy in the last 130 years. In: Visconti, G., Beniston, M., Iannorelli, E., Barba, D. (Eds.), Global Change and Protected Areas, volume 9 of Advances in Global Change Research. Springer, Netherlands, pp. 11–17.
Castelli, M., Silva, S., Vanneschi, L., 2014. A c++ framework for geometric semantic genetic programming. Genet. Program Evolvable Mach. 1–9.
Haykin, S., 1999. Neural networks: a comprehensive foundation. Prentice Hall.
Hoffmann, L., 2009. Multivariate isotonic regression and its algorithms, Wichita State University, College of Liberal Arts and Sciences. Department of Mathematics and Statistics.
Jackson, D., 2010. Promoting phenotypic diversity in genetic programming. In: PPSN 2010 11th International Conference on Parallel Problem Solving From Nature (Ed.), Lecture Notes in Computer Science vol. 6239. Springer, Krakow, Poland, pp. 472–481.
Koza, J.R., 1992. Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge, MA, USA.
Koza, J.R., 2010. Human-competitive results produced by genetic programming. Genet. Program Evolvable Mach. 11, 251–284.
Krawiec, K., Lichocki, P., 2009. Approximating geometric crossover in semantic space. GECCO'09: Proceedings of the 11th Annual conference on Genetic and evolutionary computation. ACM, Montreal, pp. 987–994.
Lance, G.N., Williams, W.T., 1967. Mixed-Data Classificatory Programs I — agglomerative systems. Aust. Comput. J. 1, 15–20.
Moraglio, A., Krawiec, K., Johnson, C.G., 2012. Geometric semantic genetic programming. In: Coello Coello, C.A., Cutello, V., Deb, K., Forrest, S., Nicosia, G., Pavone, M. (Eds.), Parallel Problem Solving from Nature, PPSN XII (part 1)Lecture Notes in Computer Science vol. 7491. Springer, pp. 21–31.
Poli, R., Langdon, W.B., Mcphee, N.F., 2008. A field guide to genetic programming.
Schölkopf, B., Smola, A., 2002. Learning with kernels: support vector machines, regularization, optimization, and beyond, adaptive computation and machine learning. MIT Press.
Seber, G., Wild, C., 2003. Nonlinear regression. Wiley Series in Probability and Statistics. Wiley.
Stadler, P., 1995. Towards a theory of landscapes. In: LÃ3pez-PeÃ±a, R., Waelbroeck, H., Capovilla, R., GarcÃ-a-Pelayo, R., Zertuche, F. (Eds.), Complex Systems and Binary NetworksLecture Notes in Physics vol. 461–461. Springer, Berlin/Heidelberg, pp. 78–163.
Vanneschi, L., 2004. Theory and practice for efficient genetic programming(Ph.D. thesis) Faculty of Sciences. University of Lausanne, Switzerland.
Vanneschi, L., Poli, R., 2012. Genetic programming â€" introduction, applications, theory and open issues. In: Rozenberg, G., BÃ¤ck, T., Kok, J. (Eds.), Handbook of Natural Computing. Springer, Berlin Heidelberg, pp. 709–739.
Vanneschi, L., Castelli, M., Manzoni, L., Silva, S., 2013. A new implementation of geometric semantic gp applied to predicting pharmacokinetic parameters. Proceedings of the 16th European conference on Genetic Programming, EuroGP'13. Springer-Verlag (To appear).
Vanneschi, L., Castelli, M., Silva, S., 2014. A survey of semantic methods in genetic programming. Genet. Program Evolvable Mach. 15, 195–214.
Weisberg, S., 2005. Applied linear regression. Wiley Series in Probability and Statistics. Wiley.