Energy and Buildings 102 (2015) 67–74

Contents lists available at ScienceDirect

Energy and Buildings

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

Prediction of energy performance of residential buildings: A genetic programming approach

Mauro Castelli [a,*], Leonardo Trujillo [b], Leonardo Vanneschi [a], Aleš Popovič [c,a]

[a] NOVA IMS, Universidade Nova de Lisboa, 1070-312 Lisboa, Portugal
[b] Instituto Tecnológico de Tijuana, Tijuana 22500, Mexico
[c] University of Ljubljana, Faculty of Economics, Kardeljeva ploščad 17, 1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 16 December 2014
Received in revised form 7 April 2015
Accepted 1 May 2015
Available online 19 May 2015

Keywords:
Energy consumption
Heating load
Cooling load
Genetic programming
Machine learning

ABSTRACT

Energy consumption has long been emphasized as an important policy issue in today's economies. In particular, the energy efficiency of residential buildings is considered a top priority of a country's energy policy. The paper proposes a genetic programming-based framework for estimating the energy performance of residential buildings. The objective is to build a model able to predict the heating load and the cooling load of residential buildings. An accurate prediction of these parameters facilitates a better control of energy consumption and, moreover, it helps choosing the energy supplier that better fits the energy needs, which is considered an important issue in the deregulated energy market. The proposed framework blends a recently developed version of genetic programming with a local search method and linear scaling. The resulting system enables us to build a model that produces an accurate estimation of both considered parameters. Extensive simulations on 768 diverse residential buildings confirm the suitability of the proposed method in predicting heating load and cooling load. In particular, the proposed method is more accurate than the existing state-of-the-art techniques.

1. Introduction

Energy consumption of buildings has received increasing interest in today's economies. As buildings represent substantial consumers of energy worldwide, with this trend increasing over the past few decades due to rising living standards, this issue has drawn considerable attention from various stakeholders (e.g. inhabitants, policy makers, industry). As reported in [1], from 1994 to 2004, building energy consumption in Europe and North America has increased at a rate of 1.5% and 1.9% per year, respectively. An even bigger increase has characterized the energy market in China, where building energy consumption has increased at more than 10% per year for the past 20 years [2]. The high level of building energy consumption and the steady increase in building energy demand require the design of energy efficient buildings and an improvement of their energy performance. Another important aspect is the effect of this continuous increase of energy consumption on the environment. According to a report of the United Nations Environment Program (UNEP) [3] buildings use about 40% of global energy, 25% of global water and 40% of global resources. In many countries, there is a real danger that, as a consequence of global warming and climate change (including growing reliance on air-conditioning) energy demand and CO_2 emissions will increase even further [4]. In an effort to reduce the impact of building energy consumption on the environment, the European Union has recently adopted a directive (European Directive 2002/91/EC [5]) requiring European countries to conform to appropriate minimum requirements regarding energy efficiency.

With regard to residential buildings, the largest part of the energy consumption is due to the use of so-called *heating, ventilation and air-conditioning* (HVAC) systems. As reported in [6], one way to alleviate the ever increasing demand for additional energy supply is to have more energy-efficient building designs with improved energy conservation properties. Ensuring that the right HVAC system is installed in a building is critical not only for providing consistent indoor comfort for families, but also for saving energy, which can be wasted by either a too-large or too-small system. Consumers and designers should have an idea of the approximate area a given piece of HVAC equipment might be expected to heat or cool under ideal conditions. For this aim, a correct estimation of the *heating load* (HL) and the *cooling load* (CL) is extremely important to achieve an efficient HVAC design for

* Corresponding author. Tel.: +351 213 828 628; fax: +351 213 872 140.
E-mail addresses: mcastelli@novaims.unl.pt (M. Castelli), leonardo.trujillo@tectijuana.edu.mx (L. Trujillo), lvanneschi@novaims.unl.pt (L. Vanneschi), apopovica@novaims.unl.pt (A. Popovič).

http://dx.doi.org/10.1016/j.enbuild.2015.05.013
0378-7788/© 2015 Elsevier B.V. All rights reserved.

buildings. The HL is the amount of heat energy that would need to be added to a space to maintain the temperature in an acceptable range and the CL is the amount of heat energy that would need to be removed from a space (cooling) to maintain the temperature in an acceptable range. The HL and CL, also known as *thermal loads*, take into account the building's construction and insulation (including floors, walls, ceilings and roof) and the building's glazing and skylights (based on size, performance, shading and overshadowing). Reliable estimations of HL and CL are of paramount importance and can have a serious impact on economy, since mistakes in these estimations may imply a waste of energy. Nevertheless, accurately predicting HL and CL is a difficult task. Most HVAC designs are still nowadays based on the personal advice of an HVAC professional, which has a subjective component and thus may be prone to errors. In such a perspective, reliable computational tools for accurately predicting HL and CL are much in demand, and this is the main motivation for the present work.

As discussed in [1], in the process of designing energy efficient buildings, it is important for architects, engineers and designers to identify which parameters will significantly influence future energy demand. After the identification of these parameters, architects and building designers usually need simple and reliable methods for rapidly estimating building energy performance, so that they can optimize their design plans. In recent years, several methods have been proposed for modeling building energy demand. The proposed methods range from traditional regression methods [7,8], to more complex machine learning techniques like artificial neural networks (ANNs) [9–11]. While the proposed techniques have shown good results, they present problems that often prevent designers and architects from using them. For instance, the performance of ANN models is strongly dependent from the design of the network's architecture, which is usually carried out in an ad-hoc and manual way.

In this paper, we propose a machine learning framework to predict the HL and the CL of a large set of residential buildings. The framework uses data related to eight input variables (relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution) to construct predictive models. The dataset used to assess the performance of the proposed method is the same as in [6]. While the methods described in [6] represent the state-of-the-art for the prediction of building energy consumption, we show that it is possible to develop a system able to achieve better results. The proposed approach combines a recently defined variant of genetic programming that integrates semantic awareness in the search process, with a local search method and a linear scaling technique.

The paper is organized as follows: Section 2 presents an overview of standard genetic programming and shows how genetic programming can be used to address a symbolic regression problem. Section 3 presents the geometric semantic genetic programming algorithm used in this work. Sections 4 and 5, respectively, describe the local search method and the linear scaling technique. Section 6 describes the dataset that has been used, the experimental settings and provides a detailed discussion of the obtained results. Furthermore, a comparison between the proposed system and other state-of-the-art methods is presented. Finally, Section 7 concludes the paper summarizing the contributions of this work.

2. Genetic programming

Genetic programming (GP) [12] is a computational method that belongs to the computational intelligence research area called evolutionary computation. GP consists in the automated learning of computer programs by means of a process inspired by biological

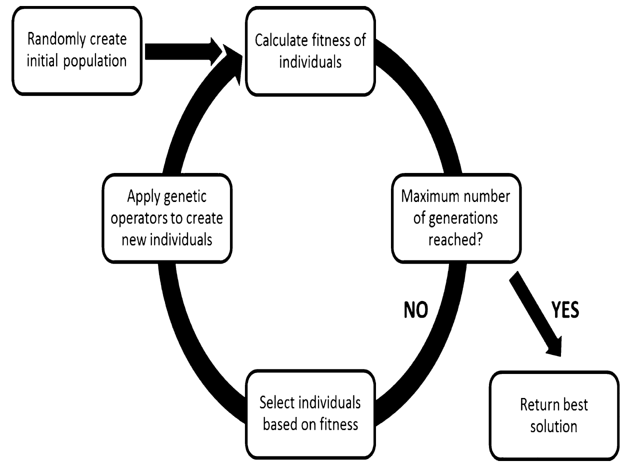

Fig. 1. The GP algorithm.

evolution. Generation by generation, GP stochastically transforms populations of programs into new, hopefully improved, populations of programs. The quality of a solution is expressed by using an objective function (also called fitness function). The search process of GP is graphically depicted in Fig. 1.

In order to transform a population into a new population of candidate solutions, GP uses particular search operators called genetic operators. Considering the common tree representation of GP individuals [12], the standard genetic operators (crossover and mutation) act on the structure of the trees that represent the candidate solutions. In other terms, standard genetic operators act on the syntax of the programs. In this paper, we use genetic operators that, differently from the standard ones, are able to act at the semantic level. The definition of semantics used in this work is the one also proposed in [13] and will be discussed in the next section.

However, to understand the differences between the genetic operators used in this work and the ones used in the standard GP algorithm, the latter are briefly described here. The standard crossover operator is traditionally used to combine the genetic material of two parents by swapping a part of one parent with a part of the other. More in detail, after choosing two individuals based on their fitness, the crossover operator selects a random subtree in each parent and swaps the selected subtrees between the two parents. The mutation operator introduces random changes in the structures of the individuals in the population. The traditional and mostly used mutation operator, called sub-tree mutation, works by randomly selecting a point in a tree, removing whatever is currently at the selected point and whatever is below the selected point and inserting a randomly generated tree at that point. This operation is controlled by a parameter that specifies the maximum size (usually measured in terms of tree depth) for the newly created subtree that is to be inserted.

2.1. Symbolic regression with genetic programming

In symbolic regression, the goal is to search for the symbolic expression $T^O : \mathbb{R}^p \rightarrow \mathbb{R}$ that best fits a particular training set $\mathbb{T} = \{(\mathbf{x}_1, t_1), \ldots, (\mathbf{x}_n, t_n)\}$ of n input/output pairs with $\mathbf{x}_i \in \mathbb{R}^p$ and $t_i \in \mathbb{R}$. The general symbolic regression problem can then be defined as

$$(T^O, \boldsymbol{\theta}^{\mathbf{O}}) \leftarrow arg \min_{T\in\mathbb{G};\boldsymbol{\theta}\in\mathbb{R}^m} f(T(\mathbf{x}_i, \boldsymbol{\theta}), t_i) \text{ with } i = 1, \ldots, p \quad (1)$$

where $\mathbb{G}$ is the solution space defined by the primitive set $\mathbb{P}$ (functions and terminals), f is the fitness function based on the distance (or error) between a program's output $T(\mathbf{x}_i, \boldsymbol{\theta})$ and the expected, or target, output t_i, and $\boldsymbol{\theta}$ is a particular parametrization of the symbolic expression T, assuming m real-valued parameters.