A machine-learning model driven by geometry, material and structural performance data in architectural design process

Sevil Yazici¹
¹Istanbul Technical University
¹sevilyazici@itu.edu.tr

Artificial Intelligence (AI), based on interpretation of data, influences various professions including architectural design today. Although research on integrating conceptual design with Machine Learning (ML) algorithms as a subset of the AI has been investigated previously, there is not a framework towards integration of architectural geometry with material properties and structural performance data towards decision making in the early-design phase. Undertaking performance simulations require significant amount of computation power and time. The aim of this research is to integrate ML algorithms into design process to achieve time efficiency and improve design results. The proposed workflow consists of three stages, including generation of the parametric model; running structural performance simulations to collect the data, and operating the ML algorithms, including Artificial Neural Network (ANN), Non-Linear Regression (NLR) and Gaussian Mixture (GM) for undertaking different tasks. The results underlined that the system generates relatively fast solutions with accuracy. Additionally, ML algorithms can assist generative design processes.

Keywords: Machine-learning, performance simulation, data-driven design, early-design phase

INTRODUCTION

Artificial Intelligence (AI), first coined by John McCarthy in 1955 (URL 1), has started to dominate various research fields including architecture and design, by aiming to respond the question of whether a machine can mimic functions of a human brain. Negroponte (1969) underlined that a machine would improve itself, if it would be asked to learn by undertaking given tasks precisely. Bayesian Models were statistical models used for decision-making, driven by results of learning from experience (Shubert, 1969). Knowledge-based tools for database design were investigated lateron, in which AI was included to provide automated assistance (Storey and Goldstein, 1993). ML and data analysis techniques enable to identify pattern on data by supporting decision-making process (Brown and Mueller, 2019). Although computational design systems are mostly
followed by structuralist approach, in which pre-defined procedures are applied, self-organizing models learn from the patterns autonomously within data (Derix and Jagannath, 2014). ML is driven by available data, which can be sourced externally or internally (Tamke and Thomsen, 2018), collected by computer-generated simulations, physical processes or experiments (Tseranidis, 2015). Precedent studies were undertaken in the past from various perspectives to incorporate AI into the early phase of architectural design processes, such as learning from standards, regulations and guidelines towards generating design alternatives (Karan and Asadi, 2019) or learning architectural style from datasets of different design projects (Strobbe et. al. 2016). A graph-based algorithm was generated additionally, where two-dimensional representation of spaces and their positioning were used by ML algorithms, of which data were driven by the BIM (As et.al, 2018).

Structural analyses, which can differ by the size and complexity of the tasks undertaken, require certain amount of computational power. (Tseranidis et. al, 2016). Although parametric models enable to explore design and objective spaces, by allowing feedback driven by performance simulations in the early phase of the design process towards decision making, they should be explicitly described by the user. Predetermination limits the parameter space of the model (Tamke et. al, 2018). Computational requirements and time needed are also high. Instead of performance simulations, ML algorithms can be used for approximation by learning from the simulation data for decision making process in a relatively fast way, as used in visual comfort of office spaces towards approximating daylight autonomy and daylight glare probability (Chatzikonstantinou and Sarıyıldız, 2016) and surrogate modelling, a class of ML algorithms, for learning tasks driven by structural simulation data to make sufficient predictions (Tseranidis, 2015).

Despite of the previous studies, more research should be undertaken towards integration of ML algorithms with design and performance properties in the early design phase. The aim of this study is to use ML to reduce the computation costs and achieve time efficiency. This research integrates ML algorithms with architectural geometry, material, and structural performance simulation data towards decision making and defines a workflow in algorithmic design environment.

**WORKFLOW OF THE PROCESS**

The workflow consists of three stages, including generation of the parametric model (1), running structural performance simulation to collect the data (2), and operating the ML algorithms (3), including Artificial Neural Network (ANN), Non-Linear Regression (NLR) and Gaussian Mixture (GM) for undertaking different tasks. Tools used in the process include Rhinoceros for geometric modelling, Grasshopper (GH) for algorithmic modelling, Karamba GH add-on for Finite Element Method (FEM) simulations, Microsoft Excel for database, and Lunchbox GH add-on for ML algorithms (Figure.1). ANN, NLR and GM algorithms are operated to predict materials, geometry and panel clusters respectively. The workflow is investigated through a case study, in which different types of ML algorithms can inform the output in the early-design phase by sharing the same parametric model.

**Generation of the Parametric Model**

While the parametric model is generated by specifying the constraints of the geometry, using Rhinoceros geometric and GH algorithmic modelling tools, the structural performance simulation engine is driven by Karamba GH add-on. The geometry is specified by sweep command generated by rail and profile curves, of which base sits into a square with 15 m span and 3.5 m surface height. It is designed as a symmetrical thin-shell structure with a surface area of 265.04 sqm. Thickness value of the shell can be set as 0.3 m or 0.5 m. The boundary conditions and materials need to be specified for the structural performance simulation, driven by the FEM. Gravity is assigned as loading and different types of materials are selected for simulations. Cross section opti-
mization for the shell is undertaken. The results are presented both as colours mapped on the geometry and numerical values.

**Running Structural Performance Simulation to Collect the Data**

By running series of FEM simulations in Karamba for different materials, the required data are collected. Currently, there are 4 main groups of materials, including steel, concrete, wood and aluminium. 21 different materials are specified under these groups with their material ID and names. As a result of the simulations, values for Mass [kg], max Displacement [m], min and max von Mises Stress [kN/cm²], Elastic Modulus (E) [kN/cm²] and Modulus of Rigidity (G) [kN/cm²] are calculated. The span [m], height [m] and surface area [m²] are considered as constants. The simulations are applied twice for all materials, by setting the thickness value of the shell as 0.3 m and 0.5 m respectively. Therefore, the database obtains 378 data items generated in Microsoft Excel, which consists of 9 columns related to the structural performance and geometry and 42 rows related to the materials.

**Operating the ML Algorithms**

Selection of the suitable ML algorithm is based on the problem that needs to be resolved. ML algorithms can operate through two different methods, as supervised or unsupervised learning. The task of the supervised learning is to learn through the training inputs and outputs and to make predictions based on approximation of the relationships among them. Unsupervised learning algorithms intent to discover structure in data by only using the inputs and without the outputs (URL 2). LunchBox is an add-on at GH and obtains an open source ML extension, which uses algorithms driven by Accord.NET, a ML framework for computer vision, computer audition, statistics applications and signal processing. Driven by the quality of the inputs provided, the algorithms are considered effective to predict sufficiently (URL 3). ANN, NLR and GM algorithms are selected for this study according to their capabilities towards undertaking different tasks. While ANN is used for classification of data driven by structural performance simulation, NLR is utilized to investigate relationships among different variables, more specifically points on the geometry. GM is suitable for clustering of components such as panels with certain degree of similarity (Figure 2). Thus, ML algorithms can be used for both design and performance data.
**Artificial Neural Network.** ANN’s are driven by the analysis of biological neural networks. They can learn a task by supervised learning, in which given examples are used to train data, in order to undertake certain tasks, or set their own associations in data environment by unsupervised learning (Derix and Jagannath, 2014). ANN’s are not programmed explicitly. Instead, they are driven by operational principles of brain, in which neurons identify the input and output layers, along with the hidden layer/s in between those. The user can specify the number of hidden layers. Weighted inputs are translated into the outputs by activation functions. ANN solver in the LunchboxML is operated by the input parameters, test, inputs, labels, neurons, alpha, iter and seed; and by the output parameter, result. While test is used as data against the training data, inputs include the training data. The algorithm is operated by two different input data sets, driven by the structural performance simulations, in which materials and surface thicknesses are the variables. The data include Elastic Modulus of materials, Mass and max Displacement for the first set; and max Displacement, max von Mises stress and Modulus of Rigidity of materials for the second one. Both sets provide material as prediction for the result. Labels are connected to the material list in the model. Number of iterations, used to teach network, ranges from 0 to 20.

**Non-linear Regression.** NLR is an algorithm that uses functions consisting of one or more variables. Relationships among different variables can be analysed, such as fitting a nonlinear curve into a point group or using a nonlinear function to generate a 3D surface from 3D points provided as input data. Solver for NLR uses sequential minimal optimization method in the LunchboxML. While the input parameters are test, inputs, output, sigma, complex and seed; the outputs parameters are result, score and error. By using this algorithm, the profile and rail curves of architectural geometry can be altered according to the optimization function. A point group driven by the profile rail curve of the shell is used as a reference initially. An approximation for a curve is generated by the given points. First the data to test against training data needs to be identified, along with the list of training inputs and outputs. Sigma of prediction and complexity of the prediction values are specified with sliders, which range from values 0 to 15 and from 1 to 25 respectively. The seed for number generator is set to 5 as constant. Result is shown with the predicted score and error, in order to interpret the relationship between the expected and predicted results. Creating different point groups and generating approximation of curves even for more complex geometries, such as free-forms, and controlling their formal properties are possible by the use of NLR al-
algorithms. Regeneration of architectural geometries would improve the design solutions.

**Gaussian Mixture.** Through GM algorithms, clusters such as panels can be generated. It can be used to analyse and group similar components, by specifying characteristics of architectural geometry. Thus, properties of individual pieces, such as the point coordinates as x-y-z, area and colour can be identified. GM solver is only used to identify different clustering options with similar panels by using pre-defined criteria and not for rationalization purposes. While inputs, components and seed are used as input parameters, the output parameters are result, likelihood and probability. The geometry is converted into 900 quadrangular panels by generating 30 U * 30 V divisions in x and y directions through the panelization algorithm. Local surface properties are evaluated at the coordinates and deconstructed into parts. While areas of individual planar closed curves are calculated, the surfaces are tested according to their planarity. Components represent number of clusters, which is controlled by a slider from 1 to 30. Using different random seed values controls the similarity of panels in one cluster, which can be identified from 0 to 10. The results are presented by graphical outputs along with likelihood and probability values. They are visualized as panel clusters, based on the approximation of panel area size and planarity.

**RESULTS AND DISCUSSION**

The purpose of this research was to integrate ML algorithms into architectural design process, in order to reduce computation time, increase efficiency and improve design solutions. Structural performance simulations were used to collect data and establish a database. Once a sufficient database could be established, using ML algorithms would enable to interpret existing data. ML algorithms were incorporated to the parametric modelling environment to undertake different tasks, including classification, regression and clustering. They were implemented to predict (1) materials based on structural performance simulation, (2) points on architectural geometry and (3) panel clusters of the shell. The results underlined that the system generates relatively fast solutions.

(1) ANN based solver was operated for two sets containing data from the structural performance simulation, including elastic modulus of the materials, mass and max displacement for the set 1; and max von mises stress, modulus of rigidity of the materials and max displacement for the set 2. Initially the system was not sufficient to predict and differentiate 21 materials. In order to overcome this problem, materials were grouped according to their main types as steel, concrete, wood and aluminium and given by a material ID as 0, 1, 2 and 3 respectively (Figure 3). By simplifying the material list connected to the labels, it was observed that the system was able to predict the materials based on simulation results accurately and rapidly. The results were achieved much faster compared to the tasks undertaken by structural performance simulations. Generalization error seems to be an important issue already in approximation models, measured by the comparison of simulation results with the predicted ones (Tseranidis, 2015). Increasing the size of the database would potentially overcome this issue and more articulated predictions would be achieved.

(2) By running the NLR solver, different geometry options were tested. The profile and rail curves of the architectural geometry were altered by reducing or increasing the slope of the arch. Thus, the height of the shell structure ranged from 3.5 m to 2.8 m. Sigma of prediction curve as degree (d) and complexity (c) parameters were variables that could be controlled towards prediction of the output curves. For instance, when the degree and complexity values were set to 15 and 10, the height of the shell became 2.8 m. By setting the complexity value as 10 and reducing the degree value to 10, the height of the shell was brought to 3.5 m height. The relationships among these two parameters informed geometrical characteristics of the surface. Additionally, transitions from complex to simpler geometries were achieved by using different point groups and adjusting their d and c values (Figure 4). NLR based ML algo-
algorithms can be used for generative design. Thus, the design space would be improved with new alternatives.

(3) By applying the GM solver to the shell structure, panel clustering conditions were identified, based on their area size and planarity parameters. Surface panels obtaining similar properties were grouped and presented by colours, probability and likelihood values. By the components value, number of clusters was altered. When components value would be increased, the likelihood value that an input belong to a cluster would be reduced. Probability gives the indication that an input belongs to a cluster and varied by alteration of the component. Initially, U and V values were set to 30 * 30 for the panelization, which affected the number of panels and their sizes. It is identified that values equal or lower than 10 * 10 for the U and V would prevent the operation of the solver, due to the insufficient number of panels. Component (c) and random seed (s) values were altered to generate iterations for panelling clusters (Figure 5). The problem was mainly due to the results, which did not represent any type of symmetry; unlike the geometrical features of the shell, which was symmetrical. However, a random distribution of panel clusters, characterized by their probability and likelihood values, were achieved. Therefore, the study didn’t provide optimal outputs for fabrication purposes, such as widely used rationalization algorithms for panels, however random transitions between clusters, which could be interpreted for design purposes, due to the random seed value.

CONCLUDING REMARKS
Parametric models are capable of exploring design, performance and optimization spaces. However,
Figure 5
GM algorithm implemented towards prediction of panel clusters based on the area size and planarity of panels.

some particular tasks, such as simulation, require significant amount of computational power and time. The aim of this research is to increase efficiency in architectural design process by incorporating the ML algorithms into one parametric design model. The outcomes of this research and future studies are the following ones:

- Using ML algorithms can be applicable towards approximation of certain tasks with increased speed and accuracy.
- By integrating ML to the GH model, feedback loops can be generated among different ML algorithms.
- ML algorithms can contribute to the generative design processes.
- The contribution of the ML-based workflow is using both design and performance data in the early design phase.
- The database should be extended in the future to improve the results in terms of accuracy.
- Incorporating other types of ML algorithms into the early design phase should be further investigated.

Acknowledgement. Thanks to Emre Toprak from ARUP Istanbul office for his contribution during our collaboration at Ozyegin University, Faculty of Architecture and Design.

REFERENCES