Paradox in Genetic Programming Modeling with Reference to Predicting Strength of Concrete

Preeti Kulkarni*, Shreenivas Londhe**, Faezehossadat Khademi *** and Sayed Mohammadmehdi Jamal****

*Associate Professor, Vishwakarma Institute of Information Technology, Pune

Preeti.kulkarni@viit.ac.in

** Professor, Vishwakarma Institute of Information Technology, Pune

shreenivas.londhe@viit.ac.in

*** Department of Civil, Architectural, and Environmental Engineering, Illinois Institute of Technology, Chicago, IL, USA.

**** Civil Engineering Department, University of Hormozgan, Hormozgan, Iran

Abstract: Concrete is a composite construction material made primarily with aggregate, cement, and water. Other cementitious material like fly ash etc. are used which makes it a highly complex material and modeling prediction of concrete strength is a difficult task. Modeling of compressive strength is done using various tools as Artificial Neural Network, Genetic Programming etc. In this work an effort is made to predict 28 day compressive strength of concrete using Genetic programming (GP). GPTIPS, an open source MATLAB based software platform for symbolic data mining (SDM) was used for model development of four models with two sets of data. The multigene genetic programming (MGGP) technique models the compressive strength of concrete by integrating the capabilities of standard genetic programming and classical regression. Data in Set1 consists of mix design parameters and water absorption parameters as inputs and Set 2 consists of mix design parameters and properties of concrete as inputs for conventional concrete and Set 3 consists of non-dimensional parameters related to recycled aggregate concrete as input parameters. The models were developed in the form of equations using Genetic Programming and it was seen that some important input parameters are excluded in the equations. However the performances of these models are noteworthy. A question thus arises about the exclusion of the parameters from the equations and its explanation through the theoretical angle.

Keywords: Genetic Programming, Recycled Concrete, Conventional concrete.

Introduction

Concrete is a widely used material and is a material with a mix of constituents like Cement, Aggregates and water. Concrete consists of aggregates which are derived from natural resources and being used for a long time and termed as Conventional concrete. Other materials which can be used in concrete as replacement to coarse or fine aggregates are recycled aggregates. These are derived from processing of construction and demolition waste. The concrete utilizing recycled aggregates is called as Recycled aggregate [1]. Compressive strength of concrete is one of the main indicators of goodness of concrete and it depends on various factors mainly the material mix quantities and properties of respective materials and concrete. A need to develop a mathematical model was seen to reduce the consumption of materials and save time using tools like Artificial Neural network (ANN), Fuzzy logic (FL), Support vector regression (SVR), Genetic Programming (GP) etc. The use of mathematical models can be very useful to understand the relative significance of various constituents used in concrete. The major limitation of the ANNs, FL and SVM is that they are not capable of providing practical prediction equations [2]. Genetic Programming (GP) works on a population of individuals, each of which represents a potential solution to a problem. A set of domain specific functions are used in conjunction with the terminal set to construct potential solutions to a given problem. For symbolic regression this could consist of a set of basic mathematical functions, while Boolean and conditional operators could be included for classification problems [3]. Genetic programming can thus be a promising tool which can model an output in form of equations which can be used easily. Application of Genetic Algorithm based neural network models for predicting the Compaction factor, VB time and Compressive strength, Tensile strength, Flexural strength and Young's modulus of High performance concrete has been studied. The weights for the network have been obtained using a genetic algorithm. GA based neural network model is able to predict the compaction factor ,vee bee time, Compressive strength, Tensile strength, Flexural strength and Young's modulus of High performance concrete satisfactorily for new problems with an accuracy of about 95% [4]. The Linear genetic programming (LGP) technique was utilized to formulate the strength capacity of Reinforced Concrete (RC) beams. The proposed design equation gives reliable estimations of the strength capacity of RC beams without stirrups. The LGP model produces better out-comes than the existing building codes, i.e., ACI 446, ACI 318, ASCE-ACI 445, CEB-FIB, CSA, DIN 1045 , EC 2, NZ, and ICC. The results of sensitivity and parametric analyses indicate that the proposed design equation is capable of capturing the underlying physical behavior of the

strength of RC beams [5]. In a study two models in gene expression programming (GEP) approach for predicting compressive strength of concretes containing rice husk ash have been developed at the age of 1, 3, 7, 14, 28, 56 and 90 days. Comparing the GEP-I and GEP-II approach models prediction with the experimental results for the testing and validation stages demonstrates a high generalization capacity of the proposed models and comparatively low error values [6]. MGGP, was employed for the analysis of the tangent modulus of elasticity of normal strength concrete and high strength concrete. Three design formulas were obtained for the prediction of modulus of elasticity of concrete. The generic MGGP model provides significantly accurate determinations of the modulus of elasticity of high strength and normal strength concrete.

The current work uses Genetic Programming as a tool to develop equations to predict 28 day compressive strength of concrete with input parameters which affect the compressive strength of concrete. Three different concrete data sets were considered with each set containining 2 models. The models were developed using GP which remain mathematically simple and are readily interpretable. Observations regarding the input parameters considered in the equations and the influence of the input parameters on output are discussed. The study is an attempt to know if a paradox exists which is due to the exclusion of some input parameters which are important as per fundamentals of concrete technology. The current paper is structured as follows: brief overview of the multigene GP algorithm is done in the next section followed by the description of data and methodology adopted for developing the equations. The developed equations are discussed further followed by results and discussions and conclusions.

Genetic Programming

Genetic programming (GP) is a biologically inspired machine learning method that evolves computer programs to perform a task. It does this by randomly generating a population of computer programs (usually represented by tree structures) and then breeding together the best performing trees to create a new population. Mimicking Darwinian evolution, this process is iterated until the population contains programs that solve the task well. GP follows the principle of survival of the fittest. GP operates on parse trees rather than on bit strings as in a GA, to approximate the equation (in symbolic form) or computer program that best describes how the output relates to input variables. For details of the same readers are referred Londhe S.N. et.al. [7]. The three genetic operations are as follows:

Reproduction: An individual is chosen from the initial population and is replicated exactly into the subsequent generation and the program which does not perform are removed. Fitness measure, selection, rank selection and tournament selection are few methods of selection from which individual are duplicated.

Cross over: it is a recombination technique, where two parent results are picked and parts of their sub-tree are exchanged in light of fact that each function holds the property 'closure' (each tree member can transform all possible argument values), every crossover operation ought to bring a legal structure. It follows the following principle: 1. Two trees are selected from the population lot, 2. One node is randomly selected from each trees, 3. Selected nodes sub trees are exchanged to bring two children of new population Mutation: it is responsible for irregular changes in a tree before it is brought into the next population. Dissimilar to crossover, it is a biogenetic and works on one single individual. Throughout mutation process either all functions or terminals are separated underneath an arbitrarily determined node and a new limb is randomly generated or a single node is exchanged with each other. Perspective to portray GP as far as the structures that experiences adaptations are: Initial structure generation, Fitness measure test which assesses the structure, Operations which change the structure, the state (memory) of the framework at each stage and the system for termating the process. The system for designating the output and parameters that control the process [8]. Linear genetic Programming (LGP) which uses a specific linear representation of computer programs. Each individual (Program) in LGP is represented by a variable-length sequence of simple C language instructions, which operate on the registers or constants from predefined sets. The function set of the system can be composed of arithmetic operations (+, - , X, /), conditional branches, and function calls (f {x, xn, sqrt, ex ,sin, cos, tan, log, ln }) [7,8]. A common aspect of GP based modeling is that the GP modeling results in fairly simpler models which could be easily interpreted for the physical significance of the input variables in making a prediction.

Multigene Symbolic Regression

A multigene individual consists of one or more genes, each of which is a "traditional" GP tree [9]. Genes are acquired incrementally by individuals in order to improve fitness (e.g. to reduce a model's sum of squared errors on a data set). The overall model is a weighted linear combination of each gene. In GPTIPS, the optimal weights for the genes are automatically obtained (using ordinary least squares to regress the genes against the output data). The resulting pseudo-linear model can capture non-linear behavior. In Multigene symbolic regression each symbolic model (and each member of the GP population) is a weighted linear combination of the outputs from a number of GP trees, where each tree may be considered to be a gene [9, 10]. Multigene symbolic regression can be implemented using GPTIPS [11]. GPTIPS was introduced that was written for the specific purpose of performing symbolic regression. GPTIPS employs a unique type of symbolic regression called multigene symbolic regression that evolves linear combinations of non-linear transformations of the input variables. When the transformations are forced to be low order (by restricting the GP tree depth) this, in contrast to "standard" symbolic regression, allows the evolution of accurate, relatively compact mathematical models of predictor – response (input – output)

data sets, even when there are a large number of input variables. Hence, it is believed that GPTIPS provides a useful, free and complementary alternative to current data analysis techniques and has a wide domain of applicability [11]. In multigene symbolic regression each symbolic model (and each member of the GP population) is a weighted linear combination of the outputs from a number of GP trees, where each tree may be considered to be a "gene". For example, the multigene model shown in figure 1 predicts an output variable using input variables $x1$, $x2$ and $x3$. This model structure contains non-linear terms (e.g. the hyperbolic tangent) but is linear in the parameters with respect to the coefficients $d0$, $d1$ and $d2$. In practice, the user specifies the maximum number of genes Gmax a model is allowed to have and the maximum tree depth Dmax any gene may have and therefore can exert control over the maximum complexity of the evolved models. In particular, we have found that enforcing stringent tree depth restrictions (i.e.maximum depths of 4 or 5 nodes) often allows the evolution of relatively compact models that are linear combinations of low order non-linear transformations of the input variables.

$$y = d_0 + d_1(0.41x_1 + \tanh(x_2 x_3)) + d_2(0.45x_3 + \mathrm{sqrt}(x_2))$$
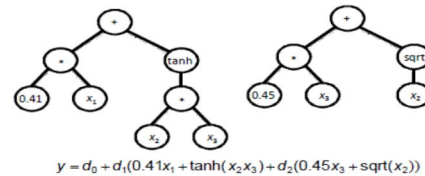
Figure. 1 Example of a multigene symbolic model

For each model, the linear coefficients are estimated from the training data using ordinary least squares techniques. Multigene GP combines the power of classical linear regression with the ability to capture non-linear behavior without needing to pre-specify the structure of the non-linear model. It was demonstrated that the multigene approach could be successfully embedded within a non-linear partial least squares algorithm [12]. In GPTIPS, the initial population is constructed by creating individuals that contain randomly generated GP trees with between 1 and Gmax genes. During the run, genes are acquired and deleted using a tree crossover operator called two point high level crossover. This allows the exchange of genes between individuals and it is used in addition to the "standard" GP recombination operators. If the ith gene in an individual is labeled Gi then a two point high level crossover is performed as in the following example. Here, the first parent individual contains the genes ($G1$ $G2$ $G3$) and the second contains the genes ($G4$ $G5$ $G6$ $G7$) where Gmax = 5. Two randomly selected crossover points are created for each individual. The genes enclosed by the crossover points are denoted by < … >.

($G1$ < **$G2$** > $G3$) ($G4$ < **$G5$ $G6$ $G7$** >)

The genes enclosed by the crossover points are then exchanged resulting in the two new individuals below.

($G1$ $G5$ $G6$ $G7$ $G3$) ($G4$ $G2$)

Two point high level crossovers allow the acquisition of new genes for both individuals but also allows genes to be removed. If an exchange of genes results in an individual containing more genes than Gmax then genes are randomly selected and deleted until the individual contains Gmax genes. In GPTIPS, standard GP subtree crossover is referred to as low level crossover. In this case, a gene is selected randomly from each parent individual, standard subtree crossover is performed and the resulting trees replace the parent trees in the otherwise unaltered individual in the next generation. GPTIPS also provides several methods of mutating trees. The user can set the relative probabilities of each of these recombinative processes. These processes are grouped into categories called events. The user can then specify the probability of crossover events, direct reproduction events and mutation events. These must sum to one. The user can also specify the probabilities of event subtypes, e.g. the probability of a two point high level crossover taking place once a crossover event has been selected, or the probability of a subtree mutation once a mutation event has been selected. However, GPTIPS provides default values for each of these probabilities so the user does not need to explicitly set them [11].

Data

The data used in the study is obtained from the experimentation carried out by the authors and data collected from literature [13-41]. The data used in the current work is divided into three sets. Each Set (Set 1, Set 2 and Set 3) consists of 2 models. Set 1 is designed with input parameters related to Recycled aggregate concrete, Set 2 with Convectional concrete and Set3 with non-dimensional parameters for recycled aggregate concrete. The data sets were designee in the said way so that the study is not limited to only one type of input parameters for a similar output. The input parameters for models in Set1 are: Content of materials in kg/m^3 for Cement (C), Natural fine aggregate (NFA), Recycle fine aggregate (RFA), Natural coarse aggregate-20mm(NCA-20), Natural coarse aggregate-10mm(NCA-10), Recycle coarse aggregate-20mm(RCA-20), Recycle coarse aggregate-10mm(RCA-10), Admixture(A), water(W). water absorption of conventional coarse aggregates (WA-NA) and water absorption of Recycled aggragtes (WA-RA). The input parameters for models in Set 2 are: content of in Kg/cm^3 of Cement (C), Sand ¾ inch (S3/4), sand 3/8 inch(S3/8), Gravel (G), w/c ratio (w/c), maximum size of sand in mm(MA), coefficient of soft sand(FM) and slump in mm (SL). The input parameters for Set 3 are ratio of natural fine aggregate to total

aggregate ratio (NFA/A), recycle fine aggregate to total aggregate ratio (RFA/A), Natural coarse aggregate-20mm to cement content (NC20/A), Natural coarse aggregate-10mm to cement content (NC10/A), Recycle coarse aggregate-20mm(RCA-20) to cement (RC20/C) , Recycle coarse aggregate-10mm(RCA-10) to cement (RC10/C), water to total materials(W/T) and replacement ratio (RR). The output for each model is 28 day compressive strength of concrete (S). The details of data used in developing the models are shown in table 1, 2 and 3. The detail of models developed in each set is shown in table 4.

Table1: Characteristics of data used for Set 1 Table2: Characteristics of data used for Set 2 Table 3: Characteristics of data used for Set 3

Sr. No	Parameters	Range of values min-max	Coorelation with output	Sr. No	Parameters	Range of values min-max	Coorelation with Output	Sr. No	Parameters	Range of values min-max	Coorelation with output
1	C	235-645	0.48	1	W/C	0.50-0.24	-0.86	1	WC	0.299-1.027	-0.585
2	NFA	0-1050	0.12	2	MA	5.12-50	0.06	2	NFA/A	0-1.566	0.152
3	RFA	0-1050	-0.18	3	G	559-1050	-0.49	3	RFA/A	0-1.221	-0.183
4	NCA-20mm	0-1508.64	0.09	4	C	243-549	0.72	4	NC20/C	0-4.726	0.007
5	NCA-10mm	0-553	0.33	5	S3/8	303-523	0.04	5	NC10/C	0-2.196	0.259
6	RCA-20mm	0-1508.64	-0.29	6	S3/4	365-693	0.09	6	RC20/C	0-5.184	-0.350
7	RCA-10mm	0-840	-0.08	7	SL	3 --6	-0.08	7	RC10/C	0-2.333	-0.113
8	W	120-358	0.02	8	FM	2.40-9.20	-0.02	8	W/T	0.055-0.174	-0.248
9	A	0-10.4	-0.31	9	S	173-394		9	RR	0-100	-0.273
10	WA-RA	1-10.60	-0.21					10	S	10.319-100.5	
11	WA-NA	0-3.56	-0.11								
12	S	10.319-100.5									

Methodology for Model Development

Concrete is a complex material and its strength characteristics depend on mix proportions of materials used and their properties [42]. To predict 28 day compressive strength of concrete and to consider mix proportions and properties of materials as input parameters, models were divided into three sets: Set 1, Set 2 and Set 3. Set1 consists of concrete made using recycled aggregates and termed as recycled aggregate concrete. Model 1 in Set 1 consists of the following input parameters: Content of materials in kg/m^3 for Cement (C), Natural fine aggregate (NFA), Recycle fine aggregate (RFA), Natural coarse aggregate-20mm (NCA-20), Natural coarse aggregate-10mm (NCA-10), Recycle coarse aggregate-20mm(RCA-20), Recycle coarse aggregate-10mm(RCA-10), Admixture(A) and water(W). Model 2 in Set consists of input parameters similar to model 1 with water absorption of Natural coarse aggregate (WA-NC) and water absorption of recycle coarse aggregate (WA-RA) as additional parameters. Set 2 consists of concrete made using conventional aggregates and thus termed as Conventional concrete. Model 1 in Set2 consists of content of in Kg/m^3 of Cement (C), Sand ¾ inch (S3/4), sand 3/8 inch (S3/8), Gravel (G) and w/c ratio (w/c). Model 2 of Set 2 consists of the parameters as said in model 1 and additional parameters as maximum size of sand in mm (MA), coefficient of soft sand(FM) and slump in mm (SL). The input parameters for Set 3 are ratio of natural fine aggregate to total aggregate ratio (NFA/A), recycle fine aggregate to total aggregate ratio (RFA/A), Natural coarse aggregate-20mm to cement content (NC20/A), Natural coarse aggregate-10mm to cement content (NC10/A), Recycle coarse aggregate-20mm(RCA-20) to cement (RC20/C) , Recycle coarse aggregate-10mm(RCA-10) to cement (RC10/C), water to total materials(W/T) and replacement ratio (RR). The details of the models are shown in the table 4. Output parameter for all the models is 28 day compressive strength of concrete(S). The total data set for each model is shown in table 3. Each model in Set1, Set2 and Set 3 were developed so that the input parameters do not remain the same and an analysis about performance of models can be done with inclusion of various parameters. For developing the models, the data is divided into 70% for training and 30% for testing.

Table 4: Details of models developed

Set No	Model No.	Input Parameters	Data numbers
1 (Recycled aggregate concrete)	1-1	C,NFA,RFA,NC-20,NC-10,RC-20,RC-10,A,W	257
	1-2	C,NFA,NC-20,NC-10,RC-20,RC-10,A,W,WA-NC, WA-RA	190
2 (Convetional concrete)	2-1	C, S3/4,S 3/8,G,W/C	150
	2-2	C, S3/4,S 3/8,G,W/C ,MA,FM,SL	150
3 (Recycled aggregate concrete parameters)	3-1	W/C,NFA/A,RFA/A,NC20/C,NC10/C,RC20/C,RC10/C,W/T	257
	3-2	W/C,NFA/A,RFA/A,NC20/C,NC10/C,RC20/C,RC10/C,W/T,RR	257

Models were developed using GPTIPS -2. Readers are referred to Searson D.P (2015) for features of GPTIPS [43]. The adopted function set to develop the GP model is given as in table 5 for each model. The parameters are selected which yielded best performance of the models. The function set used in developing the models are: Times, minus, plus, tanh, mult3, add, divide, psqroot, exp, sin, cos. These settings were based on experience with the predictive modeling of other data sets of similar size, and so they may not be optimal.

Table 5: Parameter settings for the MGGP algorithm

Model	Population	Number of generation	Selection mechanism	Max. tree depth	Probability of crossover	Probability of mutation	Max. number
1-1	900	150	tournment	4	0.84	0.14	6
1-2	700	150	tournment	4	0.84	0.14	8
2-1	300	150	tournment	4	0.84	0.14	6
2-2	1000	150	tournment	4	0.84	0.14	6
3-1	1000	150	tournment	4	0.84	0.14	8
3-2	800	150	tournment	4	0.84	0.14	6

The performance of each model is assessed using correlation coefficient (R), root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), maximum absolute error (MAbE) and sum of square errors (SSE) [43,44].

Developed Mathematical Models
As discussed earlier the models were developed using GPTIPS -2. The mathematical models developed for model 1-1 and model 1-2 are as shown in eq.1 and eq.2 below (which can be simplified further).

S = 0.673 **C** - 0.0228* **NC-20** - 0.102* **NC-10** - 0.0228* **RC-20** - 0.0228* **RC-10** + 0.912* **W** + 2.28 psqroot (**NC-10**) - 0.00307 **C*W** - 146.0 ..……..Eq. 1

S = 0.124 **C** - 1.58 10^(-4) ***NC-20** - 0.0935* **NC-10** - 0.0415 ***RC20** - 0.127* **RC-10** - 0.0882 ***A** - 0.176* **W** + 6.77 plog(**NC-10** + 2* **RC-10**) - 1.33 plog (**WA-NC** (**NC-10** + 2* **RC-10**)) - 1.58 10^(-4)* **NC-20*W** + 0.00796* **A** (**NC-10** + 2.0 ***RC-10**) + 59.9 ……….Eq. 2

The performance of all the developed models is as shown in table 6 with reference to correlation coefficient (R), root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), maximum absolute error (MAbE) and sum of square errors (SSE). The models developed for recycled aggregate concrete show a good performance and thus the model can predict the compressive strength of concrete closer to the experimental values as seen in Figure. 2.

Table 6: Performance measure of models developed

Model	R	RMSE	MSE	MAE	MAbE	SSE
1-1	0.87	7.81	60.86	5.99	26.69	4747.26
1-2	0.85	9.71	94.36	7.17	31.55	5378.72
2-1	0.94	17.45	304.76	14.18	39.21	13714.46
2-2	0.95	16.74	280.21	13.42	40.46	12609.67
3-1	0.88	7.6	57.822	6.04	20.07	4452.3
3-2	0.84	8.81	77.54	6.62	31.97	5970.2

In GPTIPS 2, a new way of analyzing the unique genes contained in a population of evolved models has been developed. This allows the user to visualize the genes in a population and to identify genes in an existing model that can be removed thus reducing model complexity whilst having only a relatively small impact on the model's predictive performance. The visualization aspect (i.e. the ability to see the gene equation and the R^2 value if the gene were removed) is important because it allows the user to rapidly make an informed choice about which model terms to remove. Often this choice is based on problem domain knowledge of the system being modelled. For example, the user might want to delete a model term such as sin(1- $x3$) because it is inconsistent with his or her knowledge about the underlying data or system. This gene-centric visualization allows users to tailor evolved models to suit their own preferences and knowledge of the modelled data. An
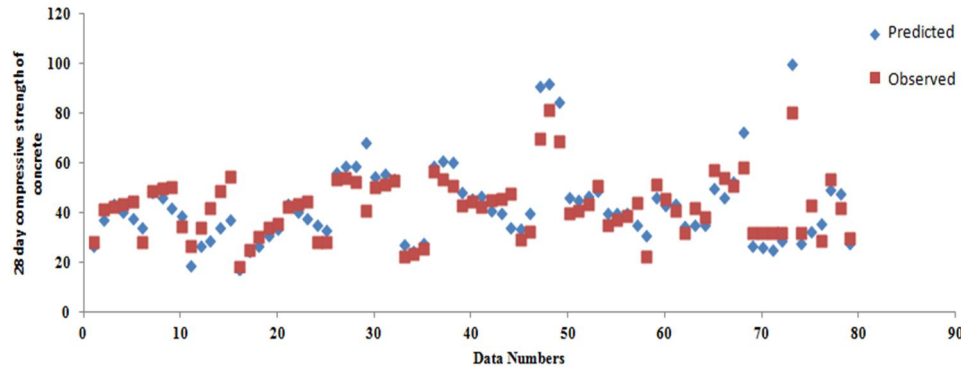
Figure. 2: Actual and Predicted values of Model 1-1

additional benefit of being able to visualize the genes in a model is that it expedites the process of human understanding of the model and intuition into which model terms account for a high degree of predictive ability and which account for lower amounts [44]. The statistical significance of each of the six genes of the derived model is visualized in Figure. 3 for model 1-1. According to the Figure. 3 the bias term is negatively very high than other genes. Figure. 3 also depicts the degree of significance of each gene evaluated using p values. As it is seen, the contribution of the genes 1, 4, 5 and 6 to explain variations in compressive strength of concrete is very high, as their relevant p values are very low. The statistical significance of the third gene (Gene 3) is lower.
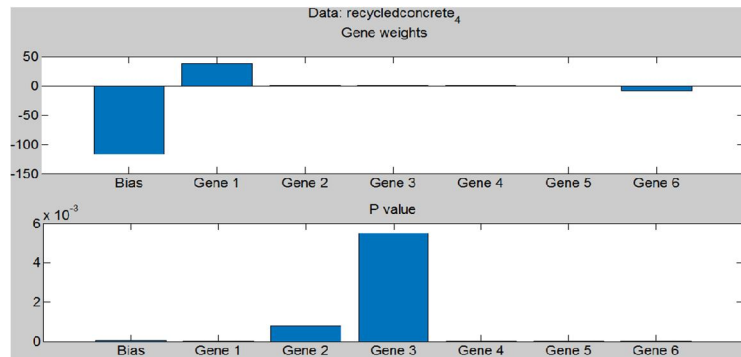

Figure. 3. Statistical properties of the evolved MGCP (on training data) for Model 1-1

The equations developed for Model 2-1 and Model 2-2 are as shown in eq. 3 and eq. 4 below (which can be simplified)

$$S= 1.58\ 10^6\ \tanh(\text{plog}(\mathbf{S3/4})) - 1.19\ 107\ \text{psqroot}(\text{plog}(\mathbf{w/c})) - 2.67\ 107\ *\mathbf{w/c} + 1.15\ 107\ \tanh(\text{plog}(\mathbf{w/c})) - 2.87\ 106\ \tanh(\mathbf{w/c}) - (1.44\ \cos(\mathbf{G}))/\text{plog}(\text{plog}(\mathbf{w/c})) + 2.99\ 107 \qquad .. \ ....\text{Eq. 3}$$

$$S = 2.99\ 107\ \tanh(\mathbf{w/c}) + (1.22\ 107\ *\mathbf{w/c}\ ^2*(\mathbf{S3/4})^2 - 3.63\ 107\ *\mathbf{S3/4^2}*\mathbf{w/c} + 1.68\ 106\ *(\mathbf{S3/4})^2 + 6.49\ 105\ *(\mathbf{S3/4})^2*(\mathbf{W/C})^5 - 2.47\ 107\ *\mathbf{w/c})/(\mathbf{S3/4})^2 - (1.71\ 105\ \tanh(\mathbf{C}))/\tanh(\mathbf{w/c}) \qquad ....\text{Eq.4}$$

The models 2-1 and 2-2 display a good performance and thus the model can be said to be predicting the strength of concrete closer to experimental values which can be visualized through the Figure. 4 (Figure. 4 for model 2-2). Figure 5 indicates the weighting coefficient of each Gene the contribution to explain the output is high in Gene 5 and followed by Gene 6.
In the similar way the equations developed in Set 3 for model 3-1 and model 3-2 are as shown in eq. 5 and eq. 6 below (which can be simplified)

$$S = 48.6\ \tanh(\mathbf{WC} + 2.0*\mathbf{NC10/C}) - 4.75*\text{psqroot}(\mathbf{RC20/C} - 2*\mathbf{RC10/C} + \mathbf{RC10/C}) + (6.1*\mathbf{NFA/A} - 6.1*\mathbf{NC10/C} + 28.1* \exp(\mathbf{WC}))/\mathbf{WC} - 83.1 \qquad ....\text{eq.5}$$

$$\mathbf{S} = 453.0\ *\mathbf{WC} - 0.0382*\mathbf{RR} + 2.45*\ \text{plog}\ (\text{plog}(\mathbf{WC} + \mathbf{NFA/A})) - 4.46*\text{plog}\ (\text{psqroot}\ (\mathbf{NC10/A})) + 5.08*\text{psqroot}\ (\text{psqroot}\ (\text{psqroot}(\mathbf{NC10/A}))) - 4.19*\exp(\ \mathbf{NC10/A}) - 4.19*\text{plog}(\mathbf{RC10/C}) - 762.0*\text{psqroot}(\mathbf{WC}) + (8.88\ 10\text{^}(-16)\ (4.54\ 1015\ *\mathbf{NFA/A} + 4.54*\ \text{psqroot}(\mathbf{NC10/C})))/\mathbf{WC} + 4.46\ *\ \mathbf{NC20/C}*\mathbf{NC10/C} + 353.0. \qquad ....\text{eq. 6}$$
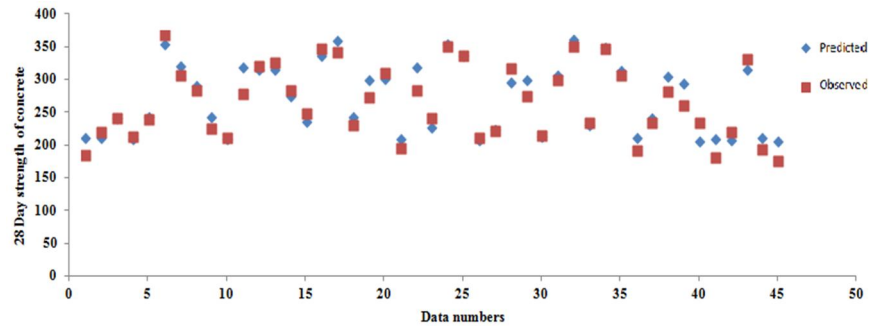
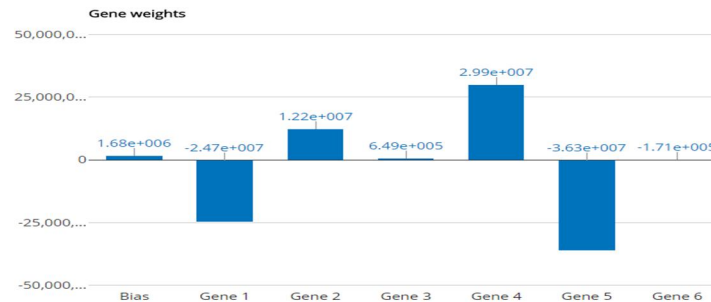Figure. 4: Actual and Predicted values of Model 2-2

Figure. 5 weighting coefficient of each Gene for model 2-2

Results and discussion

The models developed using multigene variant of the biologically inspired machine learning method of genetic programming (MGGP) as the engine drives the automatic model discovery process. Symbolic data mining is the process of extracting hidden, meaningful relationships from data in the form of symbolic equations [43]. The performances of the models developed are as shown in table 5. The performance of each of the models are satisfactory and are worthy of attention to the user community. A look at the equations show that some important parameters are not considered when the equations are developed or we can say that they are excluded from the equation developed using GP. Equation 1 shows the presence of all input parameters except NFA, RFA and A. NFA, RFA and A (Natural fine aggregate, Recycle fine aggregate and admixture contents), which are important parameters for predicting strength of concrete. Change in the proportions of these materials can show a drastic change in the strength characteristics of concrete [43]. Since the models 1-1 and 1-2 are based on recycled aggregate concrete, presence of recycled aggregate in concrete reduces the strength of concrete more than that of conventional concrete [44, 45]. Thus presence of RFA as an input parameter is important from the practical perspective. In model 1-2 the input parameters NFA (natural fine aggregate) and WA-RC (water absorption of recycle aggregates) are not considered as part of equation. However, water absorption of recycled aggregates is an important factor which increases the water demand of recycled aggregate concrete and thus further leads to reducing the strength of concrete. The correlation coefficient also shows a negative correlation of these parameters with the output. Behavior of Recycled aggregate concrete is different from behavior of conventional concrete due to presence of recycled aggregates which display properties different from conventional aggregates. Thus a next set of models were developed using parameters significant for conventional concrete. Equation 3 and equation 4 were developed with input parameters shown in table 3 for conventional concrete. Equation 3 developed using mix proportions for conventional concrete show absence of these parameters: S3/8 and Gravel. As per the domain knowledge of the characteristics of concrete, strength characteristics of concrete depend majorly on aggregate content, water cement ratio and cement content and thus in equation 3, C, S3/8 and w/c are considered and inclination towards w/c can be seen [42]. However absence of S3/8 and Gravel as input parameters in equation 3 contradicts the domain knowledge and thus an ambiguity always remains in use of these equations in practice in spite of the satisfactory performance of model. In equation. 4 the input parameters FM (soft coefficient of sand), SL (Slump) and maximum size of aggregates (MA) were added with mix proportions of conventional concrete. The equation developed displays presence of S ¾, C and w/c parameters and other parameters as FM, SL, G are not a part of the equation. In Set 3 of equations with non-dimensional parameters as input parameters, parameters-RFA/A and NC20/C are excluded during development of models. However the presences of these parameters pertaining specifically to recycled aggregate parameters are import part of the equation. The ability of GP to eliminate insignificant variables is because of the evolutionary nature of model structure optimization. By performing crossover, mutation and selection of candidate models over a number of generations GP is able

to derive the optimum model structure with the most important input variables which are relevant to the model prediction. This inturn help in developing simpler models with fewer uncertainties in the model prediction. The absence of these parameters from above equations can however makes it difficult in using these equations in practical in spite of displaying a good performance. Thus the paradox in the developed equations is the absence of these important parameters in the models developed. Thus a deep understanding of the underling process involved in development of equations is necessary. An understanding of process involved in assigning weightage to the input parameters is important. An additional attempt was made in which parameters as population, number of genes and functions were changed for model 2-2 (now termed as model 2-3). It was seen that the developed equation for model 2-3 consists of all 7 input parameters (as in model 2-2) however the performance of the model decreases to 0.93. Thus the current work develops equations for three sets of models which display a good performance and leaves the authors with the most startling thing which is its treatment towards each input parameter and ignoring few parameters which are important with reference to the domain knowledge of concrete technology. This is an interesting finding and further research in this area will help to unlock this paradox.

Conclusion

In the current paper, 3 Set of models were developed with 2 models in each set using Genetic programming in GPTIPS-2. Set 1 model consists of modeling compressive strength of recycled aggregate concrete with mix proportions and water absorption of aggregates as input parameters. Set2 consists of two models with mix proportions and properties of materials used as input parameters for conventional concrete and Set 3 consists of 3 models with non-dimensional parameters of recycled aggregate concrete. The output for each set of model was tp predict the strength of concrete. Observations of the work as follows:

1. Each model developed in the current work display a good performance. The presence of mix proportions of concrete (Recycled or conventional) as input parameters in models 1-2 and 2-1 display performance better than the models developed with additional input parameters as properties of materials.
2. The equations developed display absence of important parameters NFA, RFA, A in eq. 1, NFA ,WA-RC in eq. 2, S 3/8, G in eq. 3 and MA, SL, G and S3/4 in eq. 4. In spite of this the performance of models are note worthy.
3. Thus the paradox in the work is that the above mentioned important parameters are not considered in respective models developed which are however important from domain knowledge perspective.

A deep understanding of the underling process involved in development of equations is necessary. An understanding of process involved in assigning weightage to the input parameters is important which can be considered as the future part for research.

References

[1] M. Chakradhara Rao, S. K. Bhattacharyya ,S. V. Barai," Influence of field recycled coarse aggregate on properties of concrete", Materials and Structures, DOI 10.1617/s11527-010-9620-x.
[2] A. M. Bayazidi, G.G. Wang, H. Amir, H. Alavi and A. H. Gandomi, "Multigene Genetic Programming for Estimation of Elastic Modulus of Concrete", Mathematical Problems in Engineering, vol. 2014, 2014, pp. 1-10.
[3] Ahlam, Sharief, and A. Sheta , "Developing a Mathematical Model to Detect Diabetes Using Multigene Genetic Programming", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 3, No.10, 2014, pp.54-59.
[4] V. G. Gorphade, H. Sudarsana Rao, M. Beulah, "Development of Genetic Algorithm based Neural Network Model for Predicting Workability and Strength of High Performance Concrete", International Journal of Inventive Engineering and Sciences (IJIES), vol.2, No. 6, 2014, pp. 1-8.
[5] A. H. Gandomi, S. D. Mohammadzadeh, J. L. Pérez-Ordó˜nez and A. H. Alavid, "Linear genetic programming for shear strength prediction of reinforced concrete beams without stirrups", Applied Soft Computing, vol. 19, 2014, pp. 112–120.
[6] M. Sarıdemir, "Genetic programming approach for prediction of compressive strength of concretes containing rice husk ash", Construction and Building Materials, vol. 24, 2010, pp. 1911–1919.
[7] S. N. Londhe, "Towards predicting water levels using artificial neural networks." IEEE Xplore, Oceans 2009, 11-14 May 2009, 2009, Europe, pp. 1-6.
[8] Genetic Programming: A Novel Computing Approach in Modeling Water Flows Genetic Programming – New Approaches and Successful Applications. Chapter 9. (2012). licensee InTech
[9] J. R. Koza, "Genetic programming: on the programming of computers by means of natural selection", MIT Press, Cambridge, MA, USA, 1992.
[10] C.A. Jolla and J. Koza, "Evolving a computer program to generate random numbers using the genetic programming paradigm," in Proceedings of the Fourth International Conference on Genetic Algorithms, Morgan Kaufmann, La Jolla,CA, 1991.
[11] D. P. Searson, D. E. Leahy and M. J. Willis, "GPTIPS: An Open Source Genetic Programming Toolbox For Multigene Symbolic Regression", Proceedings of the international multi conference of engineers and computer scientists 2010, vol I, IMECS 2010, March 17-19, 2010, Hong Kong.
[12] D. P. Searson, M. J. Willis and G. A. Montague, " Co-evolution of non-linear PLS model components", Journal of Chemometrics, vol. 2, 2007, pp. 592-603.
[13] A. Ajdukiewiez , A. Kliszczewicz, "Influence of recycled aggregates on mechanical properties of HS/HPC", Cement and Concrete Composites, vol. 24,2002, pp. 269-279.

[14] J. M. Khatib, "Properties of concrete incorporating fine recycled aggregate", Cement Concrete Research,vol. 35, No. 4, 2005, pp. 763-769.

[15] T. C. Hansen and H. Narud, "Strength of recycled concrete made from crushed concrete coarse aggregate", Concrete des constr, vol. 5, No.1, 1983, pp. 79-83.

[16] T. Y. Yueh, Y. Y. Chen and C. L. Hwang," Properties of HPC with recycled aggregates", Cement and Concrete research, vol. 36,2006, pp. 943-950.

[17] J.S.Ryu," An experimental study on the effect of recycled aggregate on concrete properties",Mag .concete research,vol. 54, No.1, 2002, pp. 7-12.

[18] A. K. Padmini, K. Ramamurthy and M. S. Matthews," Relative moisture movement through recycled aggregate concrete",Cement and Concrete Research,vol. 54, No. 5, 2003, pp. 377-384.

[19] E. Dapena, P. Alaejos, A. Lobet, and D. Pérez," Effect of Recycled Sand Content on Characteristics of Mortars and Concretes", Journal of Materials in Civil Engineering, vol. 23, No. 4,2011, pp. 414-422.

[20] V. Corinaldesi, "Mechanical and elastic behavior of concretes made of recycled-concrete coarse aggregates", Construction and Building Materials, vol. 24, 2010, pp. 1616–1620.

[21] G. Fathifazl, A. Abbas, A. G. Razaqpur, O. B. Isgor, B. Fournier, and S. Foo, "New Mixture Proportioning Method for Concrete Made with Coarse Recycled Concrete Aggregate", Journal of Materials in Civil Engineering, vol. 21, No. 10, 2009, pp. 601-611.

[22] P.C. Yong, and D. C. L. Teo, "Utilization of Recycled aggregate as coarse aggregate in concrete", UNIMAS-E Journal of Civil Engineering, vol. 1, No. 1, 2009, pp.1-6.

[23] H. Yaprak, H. Y. Aruntas, I. Demir, O. Simsek and G. Durmus, " Effects of the fine recycled concrete aggregates on the concrete properties", International Journal of the Physical Sciences, vol. 6, No. 10, 2011, pp. 2455-2461.

[24] M. Chakradhara Rao , S. K. Bhattacharyya ,S. V. Barai,(2010). Influence of field recycled coarse aggregate on properties of concrete, Materials and Structures, DOI 10.1617/s11527-010-9620-x,RILEM 2010.

[25] M. Schoppe, "Shrinkage and modulus of elasticity in concrete with recycled aggregates", MS, Thesis. 2011. California Polytechnic State University, 2011.

[26] R. Kumutha and K. Vijai, "Strength of concrete incorporating aggregates recycled from demolition waste", ARPN Journal of Engineering and Applied Sciences, vol. 5, No. 5, 2010, pp. 64-71.

[27] L. Evangelista, J. de Brito, "Durability performance of concrete made with fine recycled concrete aggregates", Cement & Concrete Composites, vol. 32, 2010, pp. 9–14.

[28] J. Zega and A. A. Di Maio, " Recycled concrete made with different natural coarse aggregates exposed to high temperature", Construction and Building Materials, vol. 23,2003, pp. 2047–2052.

[29] S. Kou, "Reusing recycled aggregates in structural concrete, Phd Thesis, 2006. Hong Kong Polytechnic University.

[30] A. Konin and D. M. Kouadio, "Influence of cement content on Recycled aggregates concrete properties", Modern Applied Science, vol. 5, No. 1, 2011, pp. 23-31

[31] C. S. Poon, Z. H. Shui, L. Lam, H. Fok and S.C. Kou," Influence of moisture states of natural and recycled aggregates on the slump and compressive strength of concrete", Cement and Concrete Research, vol. 34 ,2004, pp. 31–36.

[32] D. Kotrayothar, "Recycled Aggregate concrete for structural applications", Ph.D Thesis. University of Western Sydney, Sydney 2012,

[33] S. H. Adnan, I. A. Rahman and H. M. Saman, "Recycled Aggregate as Coarse Aggregate Replacement in Concrete Mixes", ASEAN Australian Engineering Congress 2011, Kuching, Sarawak, Malaysia, 2011.

[34] S. Poon , S. C. Kou and L. Lam, "Influence of recycled aggregate on slump and bleeding of fresh concrete", Materials and Structures, vol. 40,2007, pp. 981–988.

[35] A. Domingo-Cabo , C. Lázaro , F. López-Gayarre, M.A. Serrano-López , P. Serna and J.O. Castaño-Tabares, "Creep and shrinkage of recycled aggregate concrete", Construction and Building Materials, vol. 23, 2009, pp. 2545–2553.

[36] P. Pereira , L. Evangelista and J. de Brito, "The effect of superplasticisers on the workability and compressive strength of concrete made with fine recycled concrete aggregates", Construction and Building Materials, vol. 28, 2012, pp. 722–729.

[37] M. J. Pelufo, A. Domingo, V. A. Ulloa and N. N. Vergara, "Analysis of moisture state of recycled coarse aggregate and its influence on compression strength of the concrete", Proceedings of the International Association for Shell and Spatial Structures (IASS) Symposium 2009, Valencia Evolution and Trends in Design, Analysis and Construction of Shell and Spatial Structures, 28 September – 2 October 2009, Universidad Politecnica de Valencia, Spain

[38] S. D. Agarwal, A. B.Karnnkal and P.N.Patil, "Feasibility study of use of RCA in concrete", International Journal of Research in Engineering sciences and technology, SP-II, 2011, pp. 70-75.

[39] L. R. Evangelista and J. C. de Brito, "Criteria for the use of fine recycled concrete aggregates in concrete production", International RILEM conference on Use of Recycled building materials in building and structures ,8-11 Nov 2004, Barcelona, Spain.

[40] A. Goncalves, A. Esteves and M. Vieria, "Influence of recycled concrete aggregates on concrete durability", International RILEM conference on Use of Recycled building materials in building and structures ,8-11 Nov 2004, Barcelona,Spain.

[41] M. Nikoo, F. TorabianMoghadam, and A. Sadowski, "Prediction of Concrete Compressive Strength by Evolutionary Artificial Neural Networks", Advances in Materials Science and Engineering, vol. 2015, pp. 1-8.

[42] M.S. Shetty, "Concrete Technology", 17th Ed., S. Chand and Company, New Delhi, 2005.

[43] P. Searson, " GPTIPS 2: an open-source software platform for symbolic data mining", Chapter 22 in Handbook of Genetic Programming Applications, A.H. Gandomi et al., (Eds.), Springer, New York, NY, 2015.

[44] A. T. A. Dantas, M. B. Leite and K. J. Nagahama, ".Prediction of compressive strength of concrete containing construction and demolition waste using artificial neural networks," Construction and building materials, vol. 38, 2013, pp. 717-722.

[45] N. Deshpande, S. S. Kulkarni and N. Patil, "Effectiveness of using Coarse Recycled Concrete Aggregate in Concrete", International Journal of Earth Sciences and Engineering, vol. 4, No. 6, 2011, pp 913-919.