Ascertaining the Suitability of A Worker For Safer Execution of A Construction Task: A Genetic Programming Based Modelling Approach

Chandrasekaran Sivapragasam\textsuperscript{a}, Subbiah Ajith\textsuperscript{b,}\textsuperscript{*}, Veerasimman Arumugaprabu\textsuperscript{c}

\textsuperscript{a}Senior Professor, Department of Civil Engineering, Kalasalingam Academy of Research and Education, Krishnankoil – 626126, Tamil Nadu, India.

\textsuperscript{b}Research Scholar, Department of Civil Engineering, Kalasalingam Academy of Research and Education, Krishnankoil – 626126, Tamil Nadu, India.

\textsuperscript{c}Associate Professor, Department of Mechanical Engineering, Kalasalingam Academy of Research and Education, Krishnankoil – 626126, Tamil Nadu, India.

*Corresponding Author email: ajith.s@klu.ac.in

Abstract. This research applies the Genetic Programming (GP) based modelling to understand and recommend for the kind of task in construction site a given worker with some basic traits/characteristic is suited for in order to enhance the construction site safety. The traits considered are age, experience, education level, skill set, history of accidents committed in past and knowledge of site safety. The modelling is done on data obtained from questionnaire survey from the workers in the construction sites considered in this study. The data is divided into three sets based on some specific combination of the traits/characteristics of the workers, and GP model is developed for each set. The model reveals very interesting results. All the three models recommend clearly that a worker, irrespective of how competent he is in all the other traits/characteristics, if the historical record of involvement in unsafe acts/accidents is more, he cannot be trusted with a task involving high safety risk.

Keywords: Construction, Risk, Genetic Programming, Safety

1. Introduction

Most of the activities in construction site are considered as hazardous as it involves simultaneous multiple activities such as height work, hot work, and confined space etc [1, 2]. The injury and the fatality rate of the construction workers are higher when compared to other worksites [3, 4]. Human factors are known to be the key element for injuries and accidents that occur in construction sites. There are lot of approaches which deals with the human factors such as Human Factors Analysis and Classification (HFAC), Systems-Theoretic Accident Model and Process, Reasons Swiss Cheese model and Classification of Socio-Technical Systems but these are the qualitative tools which are also used for accident investigation [5]. There are some other techniques which is used to quantify the probable human error in each task such as Human Error Assessment and Reduction Technique (HEART) [6],

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd

1
Standardized Plant Analysis Risk – Human (SPAR-H) [7], Cognitive Reliability Error Analysis Method (CREAM) [8], Technique of Human Error Rate Prediction (THERP), Success Likelihood Index (SLIM), Predictive Human Error Analysis Technique (PHEA), Systematic Human Error Reduction and Prediction Approach (SHERPA) [9] and so on. These models and techniques generally focus on accident investigation and human error quantification but do not specifically focus on the performance of the workers. Majority of the accidents can be minimized, if a right person is selected for a right task.

With globalization throughout the world, in many construction sites, workers hail from not only different states within a country but also from other countries as well bringing into the work site language, culture and other traits which varies significantly. This naturally brings in an element of sharing of culture and experiences, which over time might affect the performance of the worker. This can also be a reason for frequent injuries/accidents in the site. It would be a better strategy, therefore, allocate job to a worker based on his performance potential, which should be assessed before allocating the job to him. Obviously, this will depend on his performance in the past and will be based on traits such as age, experience, skill etc. This exercise is relatively easier when it comes to smaller construction sites were the number of workers is less, and the concerned site engineers or safety engineers can take care of this task. The issue becomes much more challenging in a relatively larger construction site with diversity of traits in the workers. If the relationship between the workers' performance is clearly identified, it is easy to focus on elements which need specific attention and which will lead to substantial improvement in the site safety.

In this study, at attempt is made to mathematically model the worker performance as a function of the traits of the workers using Genetic Programming (GP). GP falls in the category of data based evolutionary techniques and have been demonstrated for successful application in many diverse fields [10-13]. GP differs from conventional regression models in the sense that in latter, the form of the mathematical model is pre-fixed (either a linear or non-linear model) and the coefficients are obtained using the regression methods. This puts a serious limitation on the understanding of the actual nature of the process being studied. On the other hand, GP, evolves a flexible form of the model which offers a better understanding of the process. Moreover, these models can also capture the integrated effect of variables which is very difficult to obtain using conventional regression.

The remaining part of the manuscript is organized as follows. The next section discusses the overview of genetic programming which is followed by a description of the proposed methodology and case study. Results obtained are then discussed in detail and finally conclusions are drawn.

2. Genetic Programming

For numerous problems in artificial intelligence and machine learning, the best way to find the solution is through computer programs (i.e., the configuration of both primitive and terminal functions) but their size, shape and character strings has to determined earlier. It is a restrictive and difficult process in which the different configurations can be represented in a fixed character strings. To overcome these issues GP is used to solve problems of different configurations [14]. GP is one of the heuristic methods as similar to Genetic Algorithm (GA) which is grounded on the principle of Darwinian’s theory and used for symbolic regression. In this study, GP is implemented using Discipulus software.

2.1 Steps involved in using GP

There are five major steps to model the required data using GP which are as follows.

(a) The data expected to affect the model development has to be carefully collected, and pre-processed, if necessary, to make it suitable for the modelling. The total data set is divided into training, testing and validation. The model development happens through the training and testing sets, and the validation set is used for validating the quality of the model developed.
(b) A suitable primitive function has to be identified with some basic understandings of the process which includes arithmetic, mathematical, logical, exponential, logarithmic etc. For instance, if a process is not believed to be having exponential response to the input variables, the inclusion of such a function in the modelling process will not be recommended. It is not always possible to have a clear idea on which functions will influence a particular process, in which case, there has to be trial-and-error approach, and suitable justification of the finally evolved model should be carried out with additional experiments/data set not from the set used for developing the model.

(c) Fitness calculation can be done through appropriate error measures so as to get optimal models from the training process.

(d) The parameters which are used to control the run such as population size, maximum program size, subset size, number of generations, mutation rate, crossover rate etc have to be selected carefully. The cross over rate creates diversity in the solution space, while the mutation rate ensures that the process of search for optimal model is not stuck-up in local minima. A high percentage of mutation rate can, however, seriously affect the process of convergence of the training process.

(e) Once the above-mentioned parameters are selected (on a trial - and - error basis), the GP process is executed. With the progress in generation, the solution space is refined till the expected accuracy is obtained or the user decides to terminate the training process. The best evolved model is then obtained and interpreted for its physical meaning.

3. Case Study Description

In this research, building construction sites from the southern part of the Peninsular India is considered. A questionnaire survey is designed as shown in Table 1 and responses are obtained from about 100 workers in the site in elements or traits such as name, age, education, experience and involvement in accidents in the past. In addition to this, the safety engineers in the construction sites are consulted to rate the skill set and knowledge about safety procedure of the site and tasks of the workers.

| Questionnaire survey to identify the workers basic traits |
|----------------------------------------------------------|
| **Organization**                                         |
| **Name of the employee**                                 |
| **Age**                                                  |
| a) <25                                                   |
| b) 25-50                                                 |
| c) >50                                                   |
| **Education**                                            |
| a) Primary                                               |
| b) Middle                                                |
| c) Higher secondary                                     |
| **Experience**                                           |
| a) <7                                                    |
| b) 7-15                                                  |
| c) >15                                                   |
| **Type of accident**                                     |
| a) Near miss                                             |
| b) First aid                                             |
| c) Lost time injury                                      |

4. Model Development

As mentioned above, a total of 6 variables are considered for modelling the performance of the worker in terms of his ability to act safely in the assigned construction site tasks. The Safety Performance (SP)
can be functionally related to the influencing input variables age \( (V_0) \), Experience \( (V_1) \), Education \( (V_2) \), Skill \( (V_3) \), Accident \( (V_4) \), Safety Knowledge \( (V_5) \) and expressed in functional form as below:

The function sets are chosen to be only the arithmetic functions namely addition, subtraction, division and multiplication as the process of the worker safety performance is not related to more complex functions such as exponential or trigonometric functions. GP is run with an initial population size of 1024, the mutation and crossover rates of 0.83 and 0.46 respectively. The number of generations is fixed as 500 with the initial program size to be 512.

5. Results and Discussions

For each variable, it is to be noted that the value is scaled between 1, 2 and 3 (as discrete value). For example, a worker with skill set '1' means he is less competent in terms of the required skill set for a task, whereas a value of '3' indicates high competence. The output variable is also scaled in the same three levels with a value of '1' indicating poor performance and '3' indicating best performance.

In order to understand the safety performance issue more clearly, the data set is classified into 3 sets based on the combination of the input variables and each set is modelled separately. This is done by using the Kohonen classification algorithm. The average values (rounded off to next higher integer) of the variables \( V_0 \) to \( V_5 \) are tabulated in Table 2 for all the data in a given set. For instance, the set I consists of a subset of 23 data, and the average value for each variable in this subset indicates that the subset consists of workers who are all more or less well experienced, skilled and competent in knowledge of worksite safety. The potential for them to commit accident is medium (on an average for this group).

| Data Sets | No. of Data | \( V_0 \) | \( V_1 \) | \( V_2 \) | \( V_3 \) | \( V_4 \) | \( V_5 \) |
|-----------|-------------|---------|---------|---------|---------|---------|---------|
| Set I     | 23          | 2       | 3       | 1       | 3       | 2       | 3       |
| Set II    | 24          | 3       | 2       | 2       | 2       | 2       | 2       |
| Set III   | 53          | 1       | 1       | 3       | 1       | 2       | 2       |

For set II, the average characteristics of the group is found to be almost medium competency in all the 6 variables considered, whereas the average characteristics of the workers in set III is found to be less competent as far as age, experience and skill set is considered and medium competent with other variables. It is to be noted that, in all the three sets of data, the low, medium and high performers will be there. Before assigning a task to the worker, it is necessary to evaluate what kind of performance can be expected from him given the basic characteristics or information reflected by the variables \( V_0 \) to \( V_5 \).

Three different GP models are developed with for set I, 43% of the data used for training, 35% for testing and the remaining data is used for validation. For set II, 50% of the data is used for training, 29% of the data is used for testing and the remaining data is used for validation, whereas for Set III, 66% of the data is used for training, 25% for testing and the remaining data for validation. These percentages are arrived based on trial-and-error for the models giving the best performance in the performance measure. For this study, the performance measure is taken in terms of total number of matches with between predicted and actual output.

The best GP evolved models are listed as below in Equation 1, 2 and 3. It is to be noted that these models are only approximate and are obtained by simplifying the original models on sensitivity analysis to keep the form of the models simple and easily understandable.
\[ SP_I = 4.05 - 0.98V_4 \]  
\[ SP_{II} = \left[ \frac{12V_4 (1+V_4)}{V_4^3} + 6 \frac{V_4^3}{V_4} \right]^{0.25} \]  
\[ SP_{III} = V_2 \left( 1 - 0.02V_4^3 \right) \]  

It is interesting to note that the variable \( V_4 \) appears in all the equations, indicating that the safety performance of the worker depends on his historical account of accidents done. In other words, if a new worker with basic characteristics falling in the category of set I is selected to assign a task, even if he is found to be competent with respect to experience, age, skill sets, education and knowledge of site safety, if his historical account of committing accidents are more, he should not be given tasks which involves higher safety concerns. If a worker whose basic characteristics classifies him to be in set II, his performance will depend on education level, skill set, accidents committed as well as knowledge of worksite safety. In the set III, the skill set of the workers is found to be an important parameter apart from the history of accidents committed. In general, the workers belonging to set III will be assigned to low safety risk task only.

6. Conclusion

Based on the study, the following conclusions can be arrived at

(a) To ensure safety in construction site related tasks, it is not sufficient to evaluate the overall safety of the site. The suitability of the individual worker to a given task in terms of safety concern should be evaluated before assigning the task. This is expected to significantly reduce the site related unsafe acts.

(b) GP based mathematical model is developed based on the data collected through the questionnaire survey from the workers directly. The models indicate that a worker’s involvement in the site accidents or committing of unsafe acts in the past is the most influential factor in deciding his suitability for a given task, even if all his other basic traits/characteristics are good. The site engineer/safety engineer should take this factor into consideration.

(c) GP seems to be a better mathematical modelling tool compared to the conventional regression models because of its ability to evolve more flexible models, and thus offering a better explanation of the process being studied.

Reference

[1] Gunduz M, Birgonul M T and Ozdemir M 2018 Autom. Constr. 85 pp 124-134
[2] Rozenfeld O, Sacks R, Rosenfeld Y and Baum H 2010 Saf. Sci. 48(4) pp 491-498
[3] Fung I W, Tam V W, Lo T Y and Lu L L 2010 Int. J. Project. Manage. 28(6) pp 593-600
[4] Szóstak M 2019 Open Eng. 9(1) pp 312-320
[5] Zarei E, Yazdi M, Abbassi R and Khan F 2019 J. Loss Prev. Process Ind. 57 pp 142-155
[6] Williams J C and Bell J L 2015 Saf. Reliab. 35(3) pp 26-77
[7] Jahangiri M, Hoboubi N, Rostamabadi A, Keshavarzi S and Hosseini A A 2016 Saf. health work 7(1) pp 6-11
[8] Akyuz E 2015 Saf. Sci. 80 pp 77-86
[9] Kirwan B 1998 Appl. Ergon. 29(3) pp 157-177
[10] Sivapragasam C, Muttil N, Muthukumar S and Arun V M 2010 *Mar. Pollut. Bull.* **60**(10) pp 1849-1855
[11] Velay-Lizancos M, Perez-Ordoñez J L, Martinez-Lage I and Vazquez-Burgo P 2017 *Const. Build. Mater.* **144** pp 195-206
[12] Vanitha S, Sivapragasam C and Nampoothiri N V N 2016. *Int. Conf. Theoretical Comput. Sci. Discrete Math.* pp 445-452 Springer, Cham.
[13] Sharma N K, Suganya K, Sivapragasam C and Matheswaran M 2019 *IEEE Int. Conf. Intell. Tech. Control, Optim. Signal Process.* pp 1-4
[14] Koza, J R 1994 *Stat. Comput.* **4**(2) pp 87-112