

Original Research Paper

Parameter Optimization of Gradient Tree Boosting Using Dragonfly Algorithm in Crime Forecasting and Analysis

Alif Ridzuan Khairuddin, Nor Azizah Ali, Razana Alwee, Habibollah Haron and Azlan Mohd Zain

School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

Article history

Received: 05-05-2019

Revised: 15-07-2019

Accepted: 06-08-2019

Corresponding Author:
Alif Ridzuan Khairuddin
School of Computing, Faculty
of Engineering, Universiti
Teknologi Malaysia, Johor
Bahru, Malaysia
Email: alifjam1991@gmail.com

Abstract: Crime forecasting and analysis are very important in predicting future crime patterns and beneficial to the authorities in planning effective crime prevention measures. One of the challenges found in crime analysis is the crime data itself as its form, representation and distribution are varied and unpredictable. To handle such data, most researchers have been focusing on applying various Artificial Intelligence (AI) techniques as an analytical tool. Among them, Gradient Tree Boosting (GTB) is a newly emerged AI technique for forecasting especially in crime analysis. GTB possesses a unique feature among other AI techniques which is its robustness towards any data representation and distribution. Subsequently, this study would like to adopt GTB in modelling crime rates based on 8 defined crime types. Similar to other AI techniques, GTB's overall performance is heavily influenced by its input parameter configuration. To assess such a challenge, this study would like to propose a hybrid DA-GTB crime forecasting model that is equipped with a metaheuristic optimization algorithm called Dragonfly Algorithm (DA) in optimizing GTB's three main parameters namely number of trees, size of individual trees and learning rate. From the experimental result obtained, the application of DA for parameter optimization yielded a positive impact in enhancing GTB forecasting performance as it produced the smallest error compared to non-optimized GTB. This indicates that the proposed model is able to perform well using time series data with a limited and small sample size.

Keywords: Parameter Optimization, Artificial Intelligence, Gradient Tree Boosting, Metaheuristic Algorithm, Dragonfly Algorithm, Crime Forecasting, Multivariate Crime Analysis

Introduction

Crime forecasting is an analysis technique used to predict and forecast crime patterns as accurate as possible so that it forms significant insights into possible future crime trends based on past crime data. It is very helpful in analyzing and understanding the behavior of crime trends that potentially occurs in the future. Crime forecasting is an area of research that assists the authorities in enforcing early crime prevention measures (Ismail *et al.*, 2013). The advantages of crime forecasting are that it can prevent recurring crimes in specific areas or regions by analyzing the pattern of past crimes occurrence, help in allocating an appropriate resource management within a community for better police coverage and provide useful information to the authorities for planning an efficient solution in crime prevention measures.

In the last decade, the application of artificial intelligence techniques such as Artificial Neural Network (ANN), Support Vector Machine (SVM), fuzzy logic and genetic programming in crime forecasting has been extensively studied by researchers due to their capability to produce high forecasting performance accuracy. This is because artificial intelligence techniques possess some nonlinear functions which are able to detect nonlinear patterns in data (Rather *et al.*, 2017). Hence, they are able to discover a new crime pattern that never occurred in the past (Alwee, 2014). Although AI techniques are proven to be robust and able to handle various types of data structures, their performance is heavily influenced by parameter configuration. A poorly input parameter configuration leads to poor forecasting performance (Amroune *et al.*, 2018).

The ultimate objective of an artificial intelligence technique is to automatically construct an efficient model

from the data it learns without requiring tedious and time consuming human interference. An optimal parameter value is able to reduce the generalization error in most applied artificial intelligence techniques and thus improve the forecasting performance. The main difficulty in achieving such a goal is that the learning algorithms require a proper parameter configuration in order to adapt them to the particulars of a training set that fits the application's needs (Ganjisaffar *et al.*, 2011). Optimizing the parameters in an artificial intelligence technique is not an easy task because an improper parameter configuration leads to over fitting or under fitting problems that later affects the performance of the corresponding artificial intelligence technique. Thus, instead of the artificial intelligence technique attempting to predict the functional dependence between input and response variables, it will predict the training data itself (Natekin and Knoll, 2013).

From the study, it is observed that different researchers introduced different solutions to address the problems that arose in different AI techniques regarding parameter tuning. However, the approaches used have been similar where most existing works adopted a metaheuristic algorithm in optimizing the AI technique parameter (Alwee, 2014; Chen *et al.*, 2005; Hou and Li, 2009; Zhao *et al.*, 2012; Ebrahimi *et al.*, 2016; Hou *et al.*, 2018; Aadil *et al.*, 2018; Ramadas *et al.*, 2018; Xiao *et al.*, 2018). Motivated by this, the main objective of this study is to propose an improved crime forecasting model that is able to predict crime rates efficiently by properly tuning the required parameters of an AI technique using a metaheuristic algorithm.

In this study, a newly emerging AI technique in crime forecasting called Gradient Tree Boosting (GTB) is selected in developing the proposed crime forecasting model. GTB is an ensemble learning prediction model introduced by Friedman (2001). It adopts numerical optimization methods to minimize the loss function of the predictive model by integrating boosting and decision tree learning techniques. GTB's advantage is that it is capable of producing highly competitive, robust and interpretable solutions for both regression and classification problems (Friedman, 2001). In addition, the application of the boosting technique in GTB is able to avoid over fitting problems when new independent data is added (Friedman, 2001).

The proposed model is further improved by implementing a metaheuristic algorithm called the Dragonfly Algorithm (DA) in tuning and optimizing the selected parameters in GTB. DA is a recently introduced nature inspired metaheuristic optimization algorithm by Mirjalili (2016) which was inspired by the static and dynamic swarming behavior of dragonflies. DA's advantages are that it is able to improve the initial random population, converge towards the global

optimum and produce reliable results. DA is flexible as it is applicable in solving single-objective, multi-objective and discrete problems (Mirjalili, 2016).

Development of Proposed Model

Gradient Tree Boosting (GTB)

GTB develops a prediction model that is based on boosting and decision tree learning techniques. It is inspired from another statistical framework called the Adaptive Reweighting and Combining (ARC) algorithm introduced by Breiman (1997). Most decision tree learning techniques tend to grow a single large decision tree to the data which causes overfitting and high variance. To avoid such problems, the boosting technique is equipped in decision trees to minimize the variance in GTB. GTB preserves the long learner tree and grows it sequentially where it iteratively learns (boosting) and fixes the error of previous iterations (Budur *et al.*, 2015). Thus, the output result produced by the GTB has low variance and error.

The main objective of GTB is to find an estimation of the function $F(x)$ that maps all x to y values where the loss function value of $L(y, F(x))$ are minimized for each iteration m from N training set $\{y_i, x_i\}_{i=1}^N$ of known values (y, x) . In the first step of GTB, the loss function $L(y, F(x))$ is defined first. Then, the initial value of $F_0(x)$ is defined and its definition is shown in the following Equation (1):

$$F_0(x) = \arg \min_{\rho} \sum_{i=1}^N L(y_i, \rho) \quad (1)$$

$F_0(x)$ is an initial guess of successive increments ("steps" or "boosts") based on the sequence of preceding steps of $F_m(x)$. ρ is initial multiplier given by the line search of $F_m(x)$ For each successive $F_m(x)$ gradient descent boosting techniques using least square function as loss function for the next $F_{m+1}(x)$ are applied and defined as follows:

$$\gamma_i = \left[\frac{\partial L(y_i, F(x_i))}{\partial L(x_i)} \right]_{F(x)=F_m(x)}, i=1, N \quad (2)$$

The output result produces a residual called 'pseudo responses' γ_i that is later used to be fitted with the applied base or weak learner a_m . In GTB, the decision tree is applied as the base or weak learner and it is computed as shown in Equation (3):

$$a_m = \arg \min_{a, \beta} \sum_{i=1}^N [\gamma_i - \beta h(x_i; a)]^2 \quad (3)$$

In these steps, β is greedy stage-wise function that estimates $F(x)$ under the constraint that the step "direction" of $h(x; a)$ are a member of the parameterized class of functions $h(x; a)$. Next, the multiplier ρ_m is

computed given by the line search for each respective $F_m(x)$ and shown in Equation (4) below:

$$\rho_m = \arg \min_a \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \rho h(x_i; a_m)) \quad (4)$$

Finally, the $F_m(x)$ estimation is updated as an output approximation that is defined in the following Equation (5):

$$F_m(x) = F_{m-1}(x) + \rho_m h(x; a_m) \quad (5)$$

For each approximation output of $F_m(x)$, it is then stored in a set of $F_M(x)$. The trained GTB are then tested using the

new test sample data to observe its predictive performances. GTB overall framework was illustrated in Fig. 1.

Although GTB is proven to be robust and able to handle various types of data structures, like other AI techniques, its performance is also heavily influenced by parameter configuration. From the literature study conducted, there are three significant input parameters that heavily influence GTB performance namely number of trees, size of individual trees and learning rate (Saha *et al.*, 2015; Jalabert *et al.*, 2010; Guelman, 2012; Elith *et al.*, 2008; Zhang and Haghani, 2015).

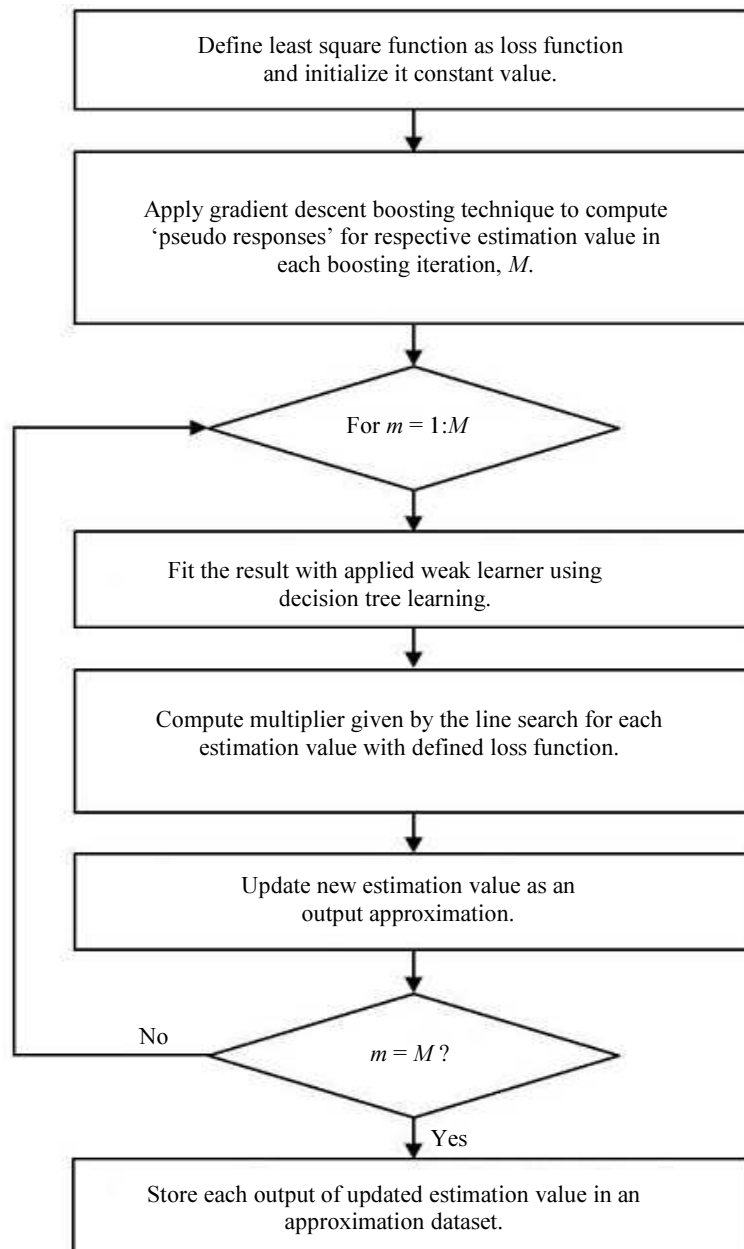


Fig. 1: GTB framework

Number of trees defines the maximum tree number used during training that controls the GTB tree complexity where the increase in number will increase the complexity. Size of individual trees defines the size of simple regression of one tree and the maximum depth of variable interactions. As for learning rate, it controls the iterations of boosting update rule where the values determine the training model convergence. These mentioned parameters are selected to be optimized in improving GTB overall performance. All three selected parameters are considered the most important as they controls GTB overall computational and performance complexity.

Dragonfly Algorithm (DA)

DA adopts swarm intelligence concepts that mimic the dragonfly's unique social interaction in navigating, migrating, food searching and avoiding enemies. Dragonfly swarming behavior is mainly based on static and dynamic features. These two swarming behavior are similar to the two main phases in the metaheuristic optimization concept; exploration (static) and exploitation (dynamic).

In static swarming, an individual or small group of dragonflies fly within a small area to search for food. The local movements and abrupt changes in the flying path of the dragonflies are the characteristic in dragonfly static swarming. The behavior of dragonflies in creating sub-swarms and moving to different places in static swarming is adopted by the exploration concept in DA. In dynamic swarm, dragonflies move in a massive swarm for migrating towards targeted places in one direction. The behavior of dragonflies moving in a bigger swarm towards targeted places with one direction is adopted by the exploitation concept in DA. The implementation of DA in this study is to identify the optimal parameter values in GTB that later improve its overall performance. Figure 2 shows the DA workflow.

Proposed Crime Model of GTB Parameter Optimization Using DA (DA-GTB)

As mentioned before, like other AI techniques, GTB is sensitive to input parameter and requires appropriate parameter tuning to optimize its performance during forecasting. Identifying the optimum parameter values of GTB that fit the crime dataset is beneficial as it is able to produce a good and reliable forecast result during crime rate forecasting. Hence, this study attempts to tackle this issue by implementing a metaheuristic algorithm called DA in identifying an optimum value of the selected parameters in GTB that later improves its overall performance.

The proposed DA-GTB model is constructed based on GTB techniques where the DA is equipped in optimizing the values of GTB selected parameters. This study is focused on solving the regression problem since this study is about forecasting crime where the main objectives are to forecast or predict crime rate values for different crime types. The proposed DA-GTB model is conducted on multivariate analysis where several factors that significantly influence crime rate are considered during forecasting and data fitting. In addition, it is focused on solving the regression problem where it is used to forecast or predict crime rate values for different crime type data (Model development is based on these crime type data). The hypothesis made is that an optimization of parameters in GTB using DA leads to a positive impact towards improving the performance of the proposed crime forecasting model for each crime type. Figure 3 shows the proposed DA-GTB crime forecasting model framework.

The framework starts with defining the list of GTB parameters (number of trees, size of individual trees and learning rate) that needs to be optimized. Once it is defined, the DA module is followed where each GTB parameter will be optimized to identify its optimum value. For the fitness function evaluation of DA for each candidate solution (possible optimum parameter value), it uses a loss value evaluation calculation. The loss value used is Weighted Mean Squared Error (WMSE) function and it is defined as follows:

$$WMSE = \sum_{i=1}^n w_i (f(x_i) - y_i)^2 \quad (6)$$

where, n is the number of data samples, $f(x_i)$ is the simulated fitted data, y_i is the actual data and is the weight vector (initial values are set to 1). In this evaluation, GTB data fitting simulation is performed to obtain the initial fitting performance. The evaluation of loss value is based on actual data and simulated fitted data. The definition made is that the lower the loss value the better the candidate solution to be selected as the best solution (optimum value). Additionally, a K-Fold cross validation is used to validate the calculated loss value.

Once an optimum value for each respective GTB parameter has been obtained, the proposed model is trained using the training dataset. The trained model is then used to forecast crime rate values using the testing dataset. In the forecasting process the model predicts the crime rate values based on the knowledge it learnt during training and the provided factors testing data. The produced forecast result is then used to evaluate the model performance. In this study, a quantitative measurement error analysis is conducted to compare the proposed hybrid DA-GTB model performance with the non-optimized GTB model.

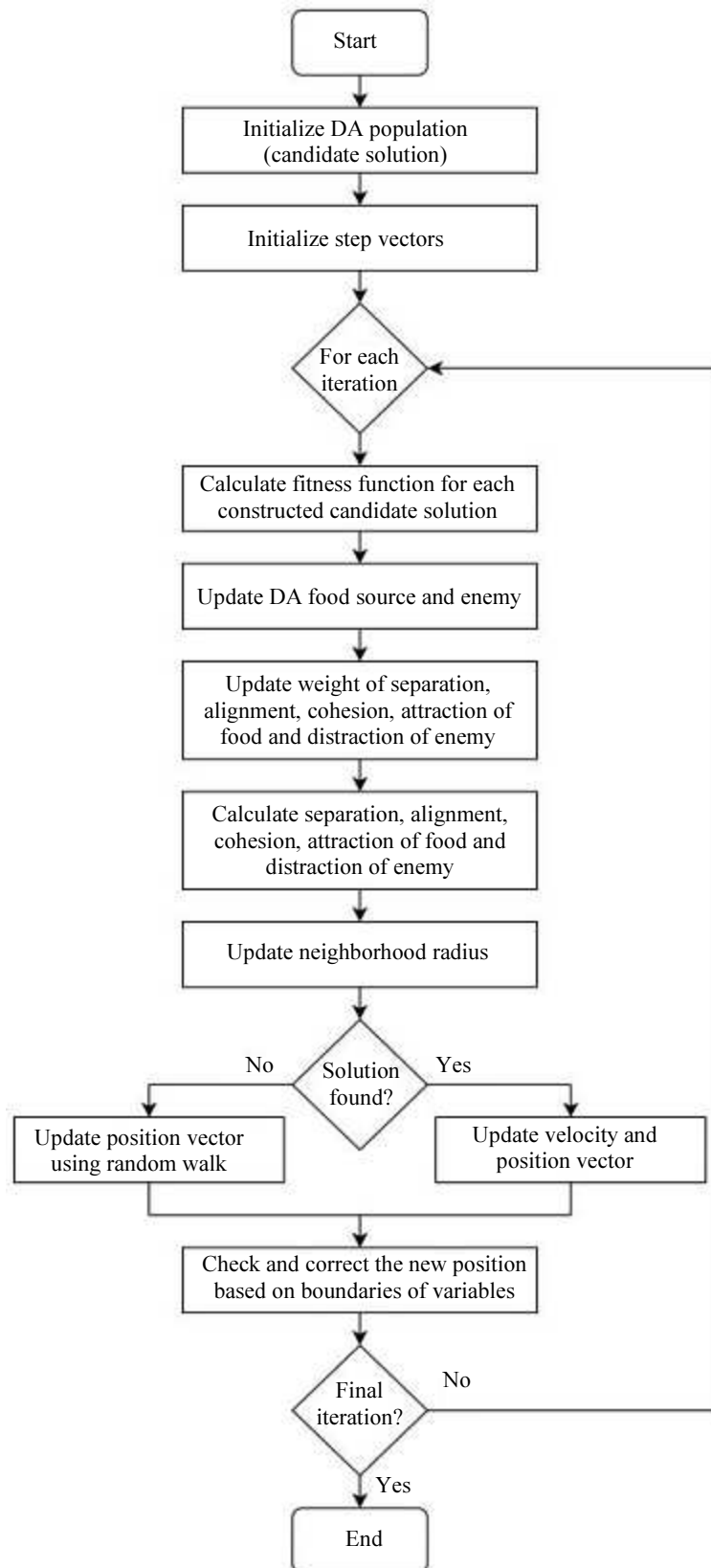


Fig. 2: Dragonfly Algorithm (DA) workflow

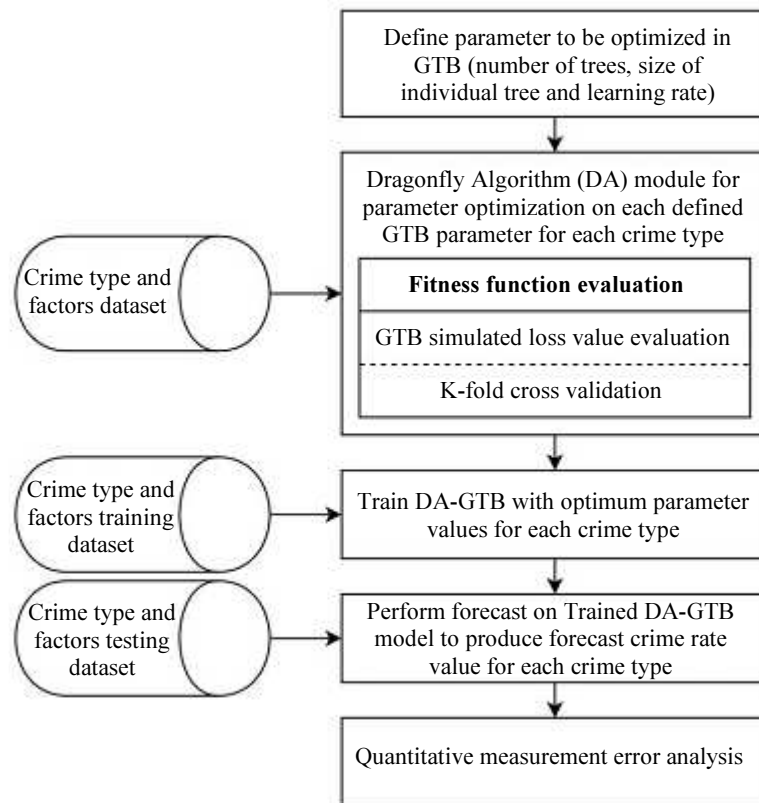


Fig. 3: Proposed DA-GTB crime forecasting model framework

Experimental Setup

The experiment is primarily conducted on the Python and Matlab platforms. In Python, Scikit-learn tools are used in modeling GTB. Scikit-learn was developed by Pedregosa *et al.* (2011) and is a Python module package that implements varieties of state-of-the-art machine learning algorithms for various problem solving solutions. It offers good flexibility in configuring the parameters and produces a consistent result. Matlab is used in developing and implementing the DA module for parameter optimization purposes. In addition, Matlab is also used for calculating the quantitative measurement error result produced from the developed crime model.

Data Definition

Two types of dataset are collected in this study; crime type dataset and factors dataset. The crime dataset is the main dataset used since the crime model is developed based on this dataset. The 8 crime type's dataset to be used in this study are murder and non-negligent manslaughter, forcible rape, robbery, aggravated assaults, burglary, larceny theft, motor vehicle theft and total crime rate for all types of crime. The crime datasets were obtained from the Uniform Crime Reporting Statistics website provided by the Federal Bureau of Investigation of the United States. For factors dataset, it

is used in developing the proposed model for multivariate analysis. The factors dataset is obtained from numerous US government agencies and other related data repository websites. Both datasets consist of annual time series data collected from 1960 to 2015 where each subset data has 56 samples. In this study, the proposed model is constructed based on these 8 crime types with their respective factors dataset. Table 1 show the dataset used in developing each crime model.

Data Preparation

The data preparation is conducted in two cases. The first case is before crime model training while the second case is after crime testing or forecasting. For the first case, the obtained raw data set (crime and factors) are preprocessed. This study implements a data normalization technique by using the feature scaling method to preprocess and transform the obtained raw time series data set of crime rate and selected factors into a dimensionless form. The normalization is to remove anomalies associated with different measurement units and scales (Alwee, 2014). The normalized data are in a scale range of between 0 and 1. The data normalization used in this study is defined as follows:

$$x' = \frac{(x_i - \min_x)}{(\max_x - \min_x)} \quad (7)$$

Table 1: Selected dataset for each crime model

Crime model	Crime dataset	Factor dataset
Murder and non-negligent manslaughter Forcible rape	US murder and non-negligent Manslaughter annual crime rate US forcible rape annual crime rate	<ul style="list-style-type: none"> • US population, male (% of total).
Robbery	US robbery annual crime rate	<ul style="list-style-type: none"> • US inflation rate. • US gross domestic product, Net exports of goods and services, export services. • US immigration statistic, total aliens apprehended. • US population, female (% of total). • US immigration statistic, lawful permanent residence status.
Aggravated assaults	US aggravated assaults annual crime rate	<ul style="list-style-type: none"> • US consumer price index, durables in city average. • US immigration statistic, total aliens apprehended. • US poverty rate, children under 18, female householder no husband present. • US population, ages 0-14 (% of total). • US tax revenue, income tax, business. • US unemployment rate, women, 16 years old and over. • US inflation rate. • US consumer sentiment index.
Burglary	US burglary annual crime rate	<ul style="list-style-type: none"> • US consumer price index, apparel in city average. • US immigration statistic, lawful permanent residence status. • US poverty rate, under 18, related children in families. • US population, male (% of total). • US tax revenue, excise taxes. • US unemployment rate, women, 20 years old and over. • US inflation rate. • US consumer sentiment index.
Larceny theft	US larceny theft annual crime rate	<ul style="list-style-type: none"> • US consumer price index, purchasing power of the consumer dollar in city average. • US immigration statistic, total aliens apprehended. • US poverty rate, with children under 18, female householder no husband present. • US unemployment rate, married women, spouse present, 16 years old and over. • US consumer sentiment index.
Motor vehicle theft	US motor vehicle theft annual crime rate	<ul style="list-style-type: none"> • US consumer price index, energy in city average. • US gross domestic product, net exports of goods and services. • US poverty rate, people between 1.00-1.25% of poverty level. • US population, male (% of total). • US tax revenue, income tax, individual. • US unemployment rate, women, 16 years old and over. • US inflation rate. • US consumer sentiment index.
Total crime rate for all types of crime	US total annual crime rate for all types of crime	<ul style="list-style-type: none"> • US consumer price index, purchasing power of the consumer dollar in city average. • US gross domestic product, gross private domestic investment, fixed investment, residential. • US immigration statistic, total aliens apprehended. • US tax revenue, excise taxes.

From Equation (7), x_j is an actual value of the selected element in the respective data series x , \max_x is maximum actual value in the respective data series x , \min_x is minimum actual value in the respective data series x and x' is the normalized value for corresponding x_j .

For the second case, after the forecasting process has been conducted, the dimensionless form of normalized forecast output values (forecasted crime rate) are subsequently transformed back into time series form of actual crime rate values through denormalization. The denormalized forecast output values are then used for the next quantitative measurement error analysis. The data denormalization ensures that the forecasted values have

the same representation with actual crime rate values to avoid unexpected errors due to different units and scales during the measurement error evaluation. The denormalization of the data is based on a logical mathematical transformation of normalization in Equation (7) and is expressed in Equation (8) below:

$$x_j = (\max_x - \min_x) + \min_x \quad (8)$$

During the experiment, the obtained data set of crime rate and selected factor is divided into two groups of training (in-sample) and testing (out-sample) data. Training data is used to train each crime model while

testing data is used to test and forecast the crime rate values based on the trained crime model. In this study, the data is divided into a ratio of 9:1 where 90% (50 samples) of the obtained data set is used for training while the remaining 10% (6 samples) is used for testing.

Initial Configuration

In the proposed DA-GTB crime model, the defined loss function used for GTB is set to Least Absolute Deviation (LAD). This loss function is based on the implementation of least square function but it attempts to identify the best solution that approximates the target data (crime data). For the other 3 selected parameters in optimizing GTB i.e., number of trees, learning rate and size of individual trees, their default non-optimized values are 100, 0.1 and 3 respectively (Zhang and Haghani, 2015). These default values are used to develop the non-optimized GTB crime model which is later used to compare its performance with the proposed DA-GTB crime model.

For DA parameter configuration, both maximum number of iterations and number of candidate solutions are set to 50. To construct the candidate solution, a value range for each selected parameter is defined first. The range serves as the boundary in constructing a set of potential optimum values for each parameter. For number of trees, the value range is set from 100 to 1000. As for the size of individual trees, it is set from 1 to 50. In learning rate, the value range is set from 0.0001 to 0.5. For cross validation configuration, 5 K-fold cross validation with 5 random sampling is used.

Evaluation Analysis

In this study, 3 types of quantitative measurement error analysis are applied to measure and compare the performance of the proposed DA-GTB model with the non-optimized GTB model. The quantitative measurement error measures the difference between forecasted crime rate and actual crime rate value for each crime type. The lower the error value, the better the performance as the model is able to forecast the crime rate value that is near to the actual crime rate value. The quantitative error measurements used are Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD) and mean absolute percentage error (MAPE). The formula to calculate the RMSE, MAD and MAPE are defined in the following Equations (9), (10) and (11) respectively:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (z_t - \bar{z}_t)^2} \quad (9)$$

$$MAD = \sum_{t=1}^n \frac{|z_t - \bar{z}_t|}{n} \quad (10)$$

$$MAPE = \sum_{t=1}^n \left| \frac{z_t - \bar{z}_t}{z_t} \right| \times \frac{100}{n} \quad (11)$$

From Equations (9), (10) and (11), n is the total number of crime rate test data used during testing process, z_t is the actual value (crime rate data for each crime type) of the selected element in the test data \bar{z}_t and is the denormalized forecasted value (forecasted crime rate for each crime type) of the selected element in the output test data.

Statistical Test

In this study, the paired sample t-test is carried out to investigate if there is a statistically significant difference between forecast output and actual crime rate data. In this study, the significance level of 95% (0.05) is considered to assess 2-tailed p -values for each model. If the p -values are larger than 0.05, this indicates that the mean difference between forecast output and actual value is not significant and thus, the model is suitable in representing the respective crime type. In contrast, if the p -values are less than 0.05, this indicates that there is a significant difference of mean between forecast output and actual value and thus, the model is deemed to be unsuitable in representing the respective crime type.

Results

The results of DA optimization conducted in finding the optimum parameter values from three selected parameters in GTB for each crime model are presented in Table 2.

Based on the optimization results shown in Table 2, it is observed that the optimum value for number of trees for murder and non-negligent manslaughter, forcible rape and motor vehicle theft falls into the range of 800 to 100 while aggravated assaults, burglary, larceny theft and total crime rate for all types of crime falls into the range of 100 to 700.

For murder and non-negligent manslaughter, forcible rape and total crime rate for all types of crime the optimum value range for size of individual trees is from 20 to 40 while robbery, aggravated assaults, burglary, larceny theft and motor vehicle theft are in the range of 1 to 3. For learning rate, the optimum value range of murder and non-negligent manslaughter, forcible rape and motor vehicle theft are from 0.05 to 0.07. In robbery, aggravated assaults and total crime rate for all types of crime, the optimum learning rate value falls towards 0.10 to 0.17 while in burglary and larceny theft, optimum values of 0.2 to 0.3 are observed.

The optimized values shown in Table 2 are then used to configure the parameters of GTB in developing the proposed DA-GTB crime models. The forecasted or predicted results for each developed crime model are then collected. Next, the quantitative measurement error is used to calculate and evaluate the performance of each proposed DA-GTB crime model. Finally, the resulting performance is compared with the non-optimized GTB

crime model to observe the significance of parameter optimization in improving GTB performance. The quantitative measurement error results obtained are presented in Table 3.

Based on the quantitative measurement error result shown in Table 3, the implementation of DA in optimizing the GTB parameter gives a positive impact towards the improvement of forecasting performance. It can be proved by observing the quantitative measurement error result that shows our proposed DA-GTB crime forecasting model outperforming the standard non-optimized GTB in all developed crime models. The result also shows the significant error minimization after optimizing the GTB parameters. This is demonstrated by its lowest RMSE, MAD and MAPE error results produced in all crime models compared to the non-optimized GTB.

To validate the statistical significance of the models' performance, paired sample t-tests are conducted and the produced results are evaluated and presented in Table 4.

The statistical test result in Table 4 shows that between forecasted and actual crime rate data, the *p-values* for DA-GTB models are larger than 0.05 for all crime types. This indicates that there is no significant difference in mean between forecast output and actual crime rate data for all developed models in all crime types. Hence, all developed DA-GTB models are considered statistically appropriate in modeling the 8 defined crime

types. For the GTB model in motor vehicle theft, the model is considered statistically inappropriate as the observed *p-value* is smaller than 0.05. In the other 7 crime types, GTB models are deemed appropriate with the observed *p-values* larger than 0.05.

To evaluate which model is more statistically appropriate, the mean values are observed and compared. In murder and non-negligent manslaughter, forcible rape, robbery, aggravated assaults, burglary and total crime rate for all types of crime, the proposed DA-GTB are statistically considered the best model as the observed mean values are smallest and near to zero compared to the GTB model. In larceny theft, the GTB model is more statistically suitable than DA-GTB. For motor vehicle theft, although GTB possessed the smallest mean, it is not statistically appropriate. Meanwhile, in DA-GTB, it is statistically appropriate and thus selected as the best model in representing motor vehicle crime.

Overall, the proposed DA-GTB model has the best performance result compared to the GTB model. The performance result is also supported and validated by the observed statistical test. Hence, an optimization made by identifying the optimal parameter values of 3 defined parameters in GTB yield positive result in improving its overall forecasting performance in forecasting 8 defined crime types.

Table 2: Optimized parameter values of GTB for each crime type

Crime Type	Optimized value		
	Number of tree	Size of individual tree	Learning rate
Murder and non-negligent manslaughter	952	32	0.0613
Forcible rape	943	35	0.0637
Robbery	988	2	0.1185
Aggravated assaults	685	1	0.1691
Burglary	568	1	0.2800
Larceny theft	553	3	0.2270
Motor vehicle theft	887	3	0.0602
Total crime rate for all types of crime	102	28	0.1566

Table 3: Quantitative measurement error result of developed model for each crime type

Crime type	Model	Quantitative measurement error		
		RMSE	MAD	MAPE
Murder and non-negligent manslaughter	GTB	0.1974	0.1638	3.5737
	DA-GTB	0.1956	0.1624	3.5451
Forcible rape	GTB	1.1218	1.0038	3.7250
	DA-GTB	1.0161	0.7961	2.9195
Robbery	GTB	21.1347	20.3792	18.8461
	DA-GTB	20.9868	20.2720	18.7471
Aggravated assaults	GTB	21.3244	20.4725	8.6541
	DA-GTB	14.7501	12.6104	5.3428
Burglary	GTB	83.7707	58.3076	10.9684
	DA-GTB	47.9732	38.2111	5.9996
Larceny theft	GTB	143.7844	118.2869	6.2926
	DA-GTB	139.4740	111.9098	6.0084
Motor vehicle theft	GTB	46.4784	45.8346	20.3380
	DA-GTB	42.5751	41.9590	18.6540
Total crime rate for all types of crime	GTB	229.6815	160.6459	5.4065
	DA-GTB	197.7586	134.4226	4.5313

Table 4: Paired sample T-test result of developed model for each crime type

Crime type	Model	Paired difference	
		Mean	<i>p-value</i>
Murder and non-negligent manslaughter	GTB	-0.0940	0.2747
	DA-GTB	-0.0939	0.2747
Forcible rape	GTB	-0.5985	0.1000
	DA-GTB	-0.2195	0.1047
Robbery	GTB	-20.3792	0.6416
	DA-GTB	-18.1054	0.0725
Aggravated assaults	GTB	-16.3058	0.0862
	DA-GTB	-11.4865	0.1670
Burglary	GTB	-58.3077	0.0823
	DA-GTB	22.2759	0.2938
Larceny theft	GTB	-74.2398	0.1960
	DA-GTB	-98.5639	0.0758
Motor vehicle theft	GTB	-45.8346	0.0122
	DA-GTB	-46.9590	0.0586
Total crime rate for all types of crime	GTB	-152.2073	0.1047
	DA-GTB	-134.4226	0.0929

Discussion

From the result analysis conducted, tuning the GTB parameters indeed produces a significant impact on its overall performance. It is observed that the proposed DA-GTB crime forecasting model outperforms the standard non-optimized GTB in all developed crime type models in terms of quantitative measurement errors. In addition, the implementation of DA in tuning the GTB parameters also yields a positive impact on its overall forecasting performance as it is able to identify an optimal parameter value in the three defined GTB parameters. Hence, the hypothesis made in this study is successfully achieved.

Conclusion

Forecasting in crime is very helpful in analyzing and understanding the behavior of crime trends that potentially occur in the future. In the last decade, it is found that researchers have shifted their interest towards the application of artificial intelligence techniques in crime forecasting due to their capability to produce high forecasting performance accuracy. Among the introduced artificial intelligence techniques, Gradient Tree Boosting (GTB) is a newly emerging technique in crime forecasting. GTB is advantageous as it is able to produce highly competitive, robust and interpretable solutions for both regression and classification problems.

GTB is a stage-wise additive framework that adopts numerical optimization methods to minimize the loss function of the predictive model which later enhances its predictive capabilities. Like other AI techniques, GTB's overall performance is heavily affected by its parameter configuration. Poorly input parameter configuration leads to poor forecasting performance in GTB. This is the motivation for this study's attempts to identify the

optimum values for three selected parameters (number of trees, size of individual trees and learning rate) which will enhance the GTB performance. The proposed solution is by implementing a metaheuristic algorithm called DA to identify the optimum values of these selected three parameters in GTB. In overall terms, the proposed hybrid crime forecasting model is based on GTB to forecast crime rate and it is also equipped with DA for parameter optimization purpose.

In general, an appropriate GTB parameter configuration is very important as it has a huge implication towards its overall performance. Thus, it is highly recommended to perform an initial parameter tuning by optimizing the required parameter values of GTB so that it yields a good and reliable forecasting result. In conclusion, the proposed hybrid DA-GTB model is able to handle and model the time series crime rate data of the 8 defined crime types in this study. Also, the proposed model is proven to be suitable in forecasting the crime rate using a small dataset since the collected data in this study has a small sample size.

Acknowledgement

This work manuscript undergoing several modifications before the final version was submitted. The corresponding author would like to thanks all co-authors in providing useful guidance and insight during manuscript writing. Authors would also like to thanks UTM Research Management Centre in providing financial support that greatly assisted the research.

Funding Information

This work was funded and supported by UTM RUG Grant Tier 2 under grant number Q.J130000.2628.15J16

from UTM Research Management Centre for School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM).

Author's Contributions

Alif Ridzuan Khairuddin: Responsible in initiating this work idea, organized the research study, perform result analysis and manuscript writing.

Nor Azizah Ali, Razana Alwee, Habibollah Haron and Azlan Mohd Zain: Responsible in supervising the corresponding author throughout research study and provide guidance during manuscript writing. Co-authors also provide final approval of the manuscript final version.

Ethics

This study is authentic work that contains original and unpublished material. The corresponding author have verified the contents and approved by other co-authors. No ethical issues involved.

References

- Aadil, F., W. Ahsan, Z.U. Rehman, P.A. Shah and S. Rho *et al.*, 2018. Clustering algorithm for Internet of Vehicles (IoV) based on Dragonfly Optimizer (CAVDO). *J. Supercomput.*, 74: 1-26.
DOI: 10.1007/s11227-018-2305-x
- Alwee, R., 2014. Swarm optimized support vector regression with autoregressive integrated moving average for modeling of crime rate. PhD Thesis, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.
- Amroune, M., T. Bouktir and I. Musirin, 2018. Power system voltage stability assessment using a hybrid approach combining dragonfly optimization algorithm and support vector regression. *Arabian J. Sci. Eng.*, 43: 3023-3036.
DOI: 10.1007/s13369-017-3046-5
- Breiman, L., 1997. Arcing the edge. Technical Report 486, Statistics Department, University of California at Berkeley.
- Budur, E., S. Lee and V.S. Kong, 2015. Structural analysis of criminal network and predicting hidden links using machine learning. *J. Comput. Sci.*, 5: 1507-05739.
- Chen, Y., B. Yang, J. Dong and A. Abraham, 2005. Time-series forecasting using flexible neural tree model. *Inform. Sci.*, 174: 219-235.
DOI: 10.1016/j.ins.2004.10.005
- Ebrahimi, E., M. Monjezi, M.R. Khalesi and D.J. Armaghani, 2016. Prediction and optimization of back-break and rock fragmentation using an artificial neural network and a bee colony algorithm. *Bull. Eng. Geol. Environ.*, 75: 27-36.
DOI: 10.1007/s10064-015-0720-2
- Elith, J., J.R. Leathwick and T. Hastie, 2008. A working guide to boosted regression trees. *J. Anim. Ecol.*, 77: 802-813. DOI: 10.1111/j.1365-2656.2008.01390.x
- Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, 29: 1189-1232.
- Ganjisaffar, Y., T. Debeauvais, S. Javanmardi, R. Caruana and C.V. Lopes, 2011. Distributed tuning of machine learning algorithms using MapReduce clusters. *Proceedings of the 3rd Workshop on Large Scale Data Mining: Theory and Applications*. Nov. 2-2, India, pp: 1-8. DOI: 10.1145/2002945.2002947
- Guelman, L., 2012. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Syst. Applic.*, 39: 3659-3667.
DOI: 10.1016/j.eswa.2011.09.058
- Hou, S. and Y. Li, 2009. Short-term fault prediction based on support vector machines with parameter optimization by evolution strategy. *Expert Syst. Applic.*, 36: 12383-12391.
DOI: 10.1016/j.eswa.2009.04.047
- Hou, Y., L. Zhao and H. Lu, 2018. Fuzzy neural network optimization and network traffic forecasting based on improved differential evolution. *Future Generat. Comput. Syst.*, 81: 425-432.
DOI: 10.1016/j.future.2017.08.041
- Ismail, S. and N. Ramli, 2013. Short-term crime forecasting in Kedah. *Proc. Soc. Behav. Sci.*, 91: 654-660. DOI: 10.1016/j.sbspro.2013.08.466
- Jalabert, S., M. Martin, J.P. Renaud, L. Boulonne and C. Jolivet *et al.*, 2010. Estimating forest soil bulk density using boosted regression modelling. *Soil Use Manage.*, 26: 516-528.
DOI: 10.1111/j.1475-2743.2010.00305.x
- Mirjalili, S., 2016. Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete and multi-objective problems. *Neural Comput. Applic.*, 27: 1053-1073.
DOI: 10.1007/s00521-015-1920-1
- Natekin, A. and A. Knoll, 2013. Gradient boosting machines, a tutorial. *Front. Neurobot.*, 7: 21-21.
DOI: 10.3389/fnbot.2013.00021.
- Pedregosa, F.G.V., G. Alexandre, M. Vincent, T. Bertrand and G. Olivier *et al.*, 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12: 2825-2830.
- Ramadas, G.C., E.M. Fernandes, A. Ramadas, A.M.A. Rocha and M.F.P. Costa, 2018. On metaheuristics for solving the parameter estimation problem in dynamic systems: A comparative study. *J. Optimizat.*, 2018: 1-21.
- Rather, A.M., V. Sastry and A. Agarwal, 2017. Stock market prediction and Portfolio selection models: A survey. *Opsearch*, 54: 1-22.
DOI: 10.1007/s12597-016-0289-y

- Saha, D., P. Alluri and A. Gan, 2015. Prioritizing highway safety manual's crash prediction variables using boosted regression trees. *Accident Anal. Prevent.*, 79: 133-144. DOI: 10.1016/j.aap.2015.03.011
- Xiao, H., W. Pei, Z. Dong, L. Kong and D. Wang, 2018. Application and comparison of metaheuristic and new metamodel based global optimization methods to the optimal operation of active distribution networks. *Energies*, 11: 85-85. DOI: 10.3390/en11010085
- Zhang, Y. and A. Haghani, 2015. A gradient boosting method to improve travel time prediction. *Transport. Res. Part C: Emerg. Technol.*, 58: 308-324. DOI: 10.1016/j.trc.2015.02.019
- Zhao, J., W. Wang, W. Pedrycz and X. Tian, 2012. Online parameter optimization-based prediction for converter gas system by parallel strategies. *IEEE Trans. Control Syst. Technol.*, 20: 835-845. DOI: 10.1109/TCST.2011.2134098