Improvement of Criminal Identification by Smart Optimization Method

[1]*Reem* Razzaq Abdul Hussein*, [2]Dr.Muayad* Sadik Croock*. and [3]Dr Salih* Mahdi Al-Qaraawi

[1]Informatics Institute of Postgraduate Studies, University of Information Technology and Communications. Baghdad, Iraq
[2]Computer Engineering Department, University of Technology, Baghdad, Iraq
[3]Computer Engineering Department, University of Technology, Baghdad, Iraq

Abstract. Data-mining methods, which can be optimized via different methods, are applied in crime detection. This work, the decision tree algorithm is used for classifying and optimizing its structure with the smart method. This method is applied to two datasets: Iraq and India criminals. The goal of the proposed method is to identify criminals using a mining method based on smart search. This contribution helps in the acquisition of better results than those provided by traditional mining methods via controlling the size of the tree through decreasing leaf size.

1. Introduction

For data mining, classification techniques are widely used in e-government, especially in the criminology field [1]. They help police departments to predicate criminals and information about crime locations. These techniques also can be optimized with latent models [2] or using *hyperparameters* that allow parameters to be tuned. A machine learning (ML) model may require different constraints, model selection, and learning rates to make generalizations for various data patterns. These measures are called hyperparameters, which must be tuned for the model to solve ML optimally [3]; previous researchers have used hyperparameters in different strategies; several researchers have used grid search, which is commonly known as a brute force or exhaustive search; problems, such as high dimensionality and parallelization, occur because the hyperparameter settings it generates evaluates independently for each other [4]; other researchers have used random search, rather than brute force search; in our work, we depend on Bayesian optimization (BO) as a sequential model-based optimization algorithm, which depends on the outcomes of the previous iteration to improve the sampling method for the subsequent experiment. others used BO for hyperparameters to generalize the Gaussian procedure. Support vector machines, random forests, and AdaBoost have been applied to identify the highly relevant classifier of hyperparameters and which tends to be improved [5]; a crime dataset has been utilized to find the probability distribution for predicting crimes; that used hyperparameter [6][7].In this work, we attempt to overcome the problems in previous studies using BO for k_fold DT, with less time than other previous methods because the BO method makes fewer evaluations. The main contribution is using the smart search method combined with DT to obtain an accurate model, in which the leaf size is decreased and the pruning level can be improved to reduce the cross-validation loss of DT. Two datasets are adopted to test the proposed algorithm and to consider them in comparison evaluation. The proposed method is also offering help to employees of Iraq National Identifiers in criminal detection.

2. DT Algorithms with Hyperparameter Optimization

Classical DT algorithms are widely used in data-mining methods as they utilize a vast amount of data for classification to construct a model. These algorithms include induction of DTs (ID3), C4.5, chi-squared automatic interaction detectors, and classification and regression trees [1][8], which is the adopted type in this work. The basic algorithm used to build DT is the greedy algorithm, which can perform recursion from top to bottom to build a DT. A DT is a tree where each node represents a feature, each branch represents a decision (rule), and each leaf represents an outcome. DT algorithms, which split nodes, are depending on *entropy*. Features (attributes) that lie in the lower levels of trees have low importance. The DT model generally includes the following types of nodes:

- **One root:** represents the (top) node, which is the most important feature.
- **Internal nodes:** represent an attribute that is branched from parent and generates leaf or sub-internal nodes.
- **Leaf nodes (decision):** represent a class label. The path from the root node to every decision indicates the sequence. The complexity of the tree depends on the number of leaves [9].

The tree size is difficult to be controlled during the construction of a DT. Most improved techniques adopt pruning methods [5] to solve the problem of overfitting, which can lead to the entire process of building DT models in two steps, namely, modeling and pruning. A concise DT saves considerable time and yields good outcomes.K-fold estimates the performance of a learning method when the dataset size is small or medium. Optimization generally locates a point that minimizes a real-valued function named objective function. BO internally maintains a Gaussian process model of

[*] Corresponding author: reemrazzak@yahoo.com,reem@uoitc.edu.iq

the objective function and uses objective function evaluations to train the model. One of the strategies that involve the use of BO is the adoption of an acquisition function, which the algorithm uses to specify the next point for evaluation. The acquisition function can Balance sampling at points that have low-modeled objective functions and Find areas that are not modeled clearly. BO is part of statistics and machine learning (ML) because it is well suited for optimizing the hyperparameters of classification algorithms. A hyperparameter is an internal parameter of a classifier method, such as the box constraint of DT or a support vector machine, or maybe the learning rate of a robust classification ensemble. The goal of these parameters can highly affect the performance of a classifier; BO uses a fit function [9]. For Model selection, depend on Hyper-parameters, features, train a classifier like a decision tree, using a dataset, this process needs to choose the best feature set and hyperparameters by applying K fold DT.

3. Proposed Method

Here, we need to the prediction of crimes, The first dataset is collected Iraqi Dataset collected from is collected from the Iraq Ministry of Interior website available at and Social Media Facebook of Iraq Ministry of Interior, and the second dataset about Crime in India by Rajanand Llangollen, the number of states is 34 and from 2001 to 2012 had been considered. The proposed algorithm can be explained as a flowchart as shown in Fig .1.
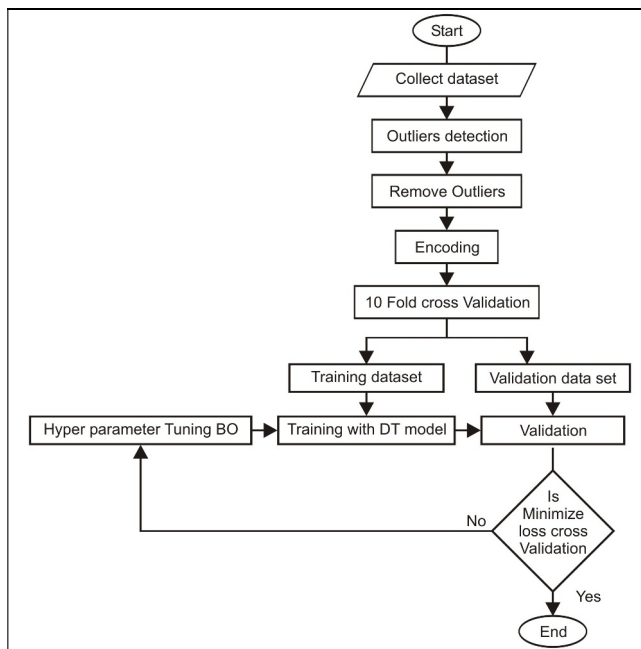

Fig. 1. The Proposed Model

The work steps of the proposed algorithm are outlined as follows:
1. The input of model include :The Iraq dataset consists of features such as as{Age, Gender, ID, Crime types, locations, Gang, longitude, latitude }the type of features is categorization whereas the features of Indian data set is included {states, murder, attempted to murder ,…, thieft} , we convert the continuous data to category which is called encoding .
2. Preprocessor: The outlier from the second dataset is removed, depending on the median absolute deviation.
3. A model is constructed by applying a binary DT with multiple classifier algorithms (ID3) to both datasets, with k_fold (10). The datasets are divided into training, validation datasets.
4. ID3 selects the best feature in the root and eliminates the worst features. Its base is work entropy, as illustrated in Section (I).
5. For smart optimization, BO hypermeter is fed the DT model, then it tested by validation dataset, which is a test whether it minimizes the loss cross-validation errors, or not.
6. The BO the minimize error of the model depends on the factors such as a number of decision tree split (branch nodes) by indicating the maximum of splitting, The minimum number of leaf nodes of DT. Split criterion, is ID3 BO, used to control the behavior of an algorithm, are used. This optimization minimizes the number of leaves in a tree and the tree level to improve accuracy.

7. The output of model DT, which determine the Nid of criminals for Iraq data set and kidnapping in second data set to get more accurate results.

4. Result and Implementation

In our work, we implement DT on two datasets and then perform optimization by using the BO hyperparameters of cross-validation, which allows us to acquire a model that provides accurate results. Table.1 shows a comparison of the two datasets in terms of different parameters.

Table 1. Comparison of datasets

	Dataset 1	Dataset 2
No. of iterations	30	30
Total elapsed time	16.5749s	19.6748s
Total objective function evaluation time	2.1775	2.9177
Best observed feasible point	MinLeafSize=1	MinLeafSize=14
Observed objective function value	0.02	0.057685
Estimated objective function value	0.019988	0.058786
Function evaluation time	0.079059	0.09567
Best estimated feasible point (according to models)	MinLeafSize=3	MinLeafSize=14
The accuracy of DT before optimization Approximately	96%	93.6%
The accuracy of DT after optimization Approximately	98%	93.8%

The accuracy of DT before optimization of the first dataset is approximately 96%, while it is ~93.6% for the second dataset. The accuracy of DT after optimization of the first dataset is improved to be approximately 98%. For the second dataset, the accuracy is also improved to be ~93.8 % after applying the proposed optimization method. The best estimated feasible point is obtained when the minimum leaf size is equal to three for both datasets, In each implementation. the outcomes may be changed. The reason behind this change is the random selection of the data partition. Fig. 2 illustrates the minimum leaf size versus the estimated objective function value.
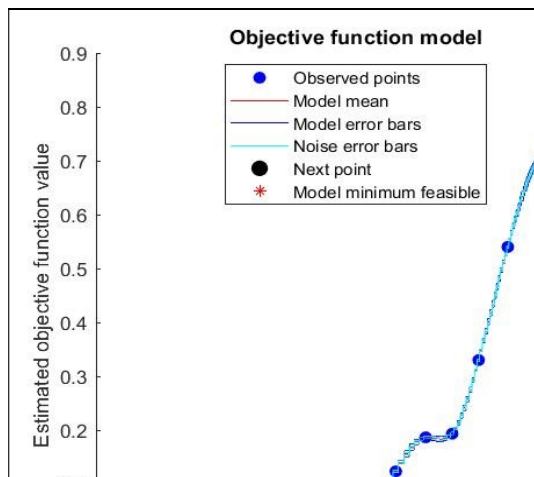

Fig 2.Objective function model

Few objective functions and evaluations are shown in Fig. 3 which are related to the first datasets. The estimation function with minimum objective function in 0.02
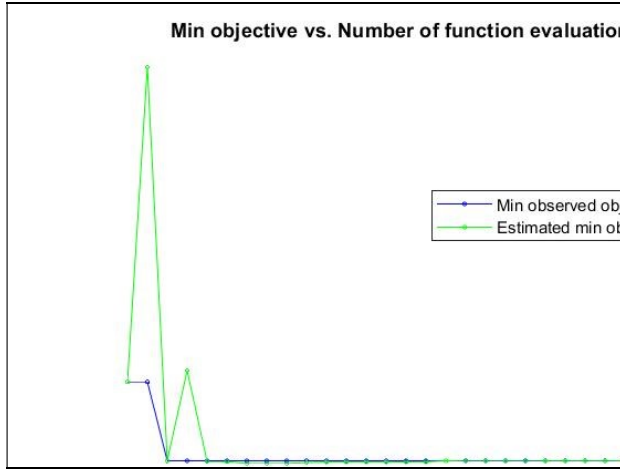

Fig. 3: Min Objective vs. no. of Function Evaluations

The corresponding accuracy of the proposed DT for the second dataset are shown in Fig.4 is equal to %98. The tree predicts classifications based on many predictors and label (id). Prediction starts at the root node, represented by a triangle (Δ). The first decision is on whether national identifier N_{id} is smaller than 12.5. If so, then the left branch of the tree is followed. a gang, which is the second level of the tree. The subset is branched from the right branch of the root level and continues until no more classification needs to be performed.
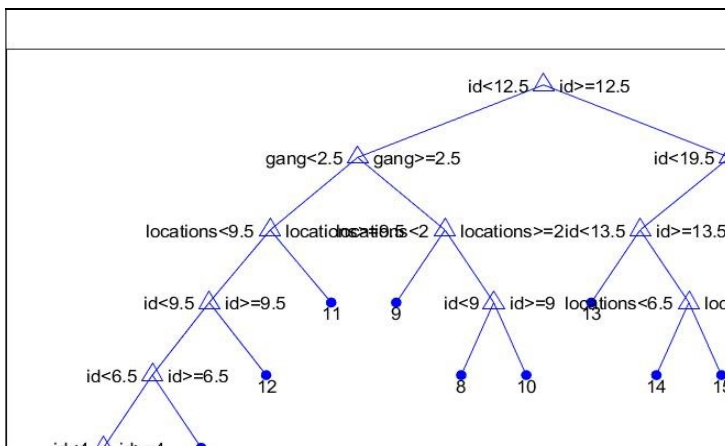

Fig .5 DT of Iraq dataset after optimizations

6. Conclusion

Criminal detection is an important issue in the prediction process. Nid can contribute to facilitating police control to identify criminals. In our work, we use a DT for modeling the proposed crime prediction algorithm. In this algorithm, the accuracy depends on labels and predictors. BO with ID3 algorithms is utilized to build a model with a small number of leaf and a minimal number of splits to obtain accurate results and minimal DT rules. We compare the performance of the proposed using two datasets. The accuracy of the first dataset is improved to approximately ~98%, whereas that of the second dataset is enhanced to ~93.8 % after optimization.in the future, we can enhance the accuracy of DT, by using machine learning   strategy to improve make the decision process

References

1]: *Reem Razzaq Abdul Hussein, Dr. Salih Mahdi Al-Qaraawi, Dr.Muayad Sadik Croock,*" Role of Data Mining in E-Government Framework". Iraqi Journal for Computers and Informatics Vol. 44 Issue 1,(2018).

[2]: *Reem Razzaq Abdul Hussein Dr.Muayad Sadik Croock, Dr. Salih Mahdi Al-Qaraawi*, "Developed Crime Location Predication Using Latten Markov Model", Journal of Theoretical and Applied Information Technology, Vol. 96. No 1, (2019.)

[3]: Claesen, Marc, and Bart De Moor. "Hyperparameter search in machine learning." arXiv preprint arXiv:1502.02127 (2015).

[4]: Bergstra, James, and Yoshua Bengio. "Random search for hyper-parameter optimization." Journal of Machine Learning Research 13, no. Feb (2012): 281-305.

[5]: Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams. "Practical Bayesian optimization of machine learning algorithms." Advances in neural information processing systems. (2012).

[6]: Perez, Luis, and Alex Wang. "Gaussian Processes for Crime Prediction.",(2014).

[7]: Dong, Kun, David Eriksson, Hannes Nickisch, David Bindel, and Andrew G. Wilson. "Scalable log determinants for Gaussian process kernel learning." In *Advances in Neural Information Processing Systems*, pp. 6327-6337.(2017).

[8] Liu, Dong-sheng, and Shu-Jiang Fan. "A modified decision tree algorithm based on genetic algorithm for mobile user classification problem." *The Scientific World Journal* 2014 (2014).

[9]: Wang, Yingying, et al. "Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree." *Algorithms,* Vol. 10, No. 4,(2017).

[10]: Yi-bin, Li, Wang Ying-Ying, and Rong Xue-wen. "Improvement of ID3 algorithm based on simplified information entropy and coordination degree." IEEE *Chinese Automation Congress (CAC2017*, (2017).