HYPERPARAMETER OPTIMIZATION OF DATA MINING ALGORITHMS ON CAR EVALUATION DATASET

Ömmu Gülsüm UZUT

Res. Asst. Dept. of Computer Engineering, Muş Alparslan University, Muş/Turkey

https://orcid.org/0000-0002-6602-772X

Selim BUYRUKOĞLU

Assist Prof. Dr. Dept. of Computer Engineering, Çankırı Karatekin University

Çankırı/Turkey, https://orcid.org/0000-0001-7844-3168

Abstract

Data mining is the process of obtaining valuable data from large-scale data. Several algorithms are used for revealing the relationships between data and making accurate predictions. There are several cases that may affect the performance of these algorithms. One of these is choosing most suitable hyper parameters. To optimize these parameters provide us to improve algorithm results. In our project, we optimized hyper parameters of different data mining algorithms on car evaluation dataset for improving classification accuracy. Hyper parameter optimization was performed on support vector machines, k-nearest neighbor, random forest, multi-layer perceptron and gradient boosting algorithms. Results of hyper parameter values and default parameter values were compared. The calculations show that gradient boosting with hyper parameter optimization method produces best prediction of the car evaluation by 99.42%.

Keywords: Data Mining, Gradient Boosting, Classification.

1. INTRODUCTION

Data mining is the process of obtaining valuable data from large-scale data. It becomes possible to reveal the relationships between the data and make accurate predictions. Technology companies and institutions are working on big data. Being able to extract useful information from a large chunk of data is a real laborious task. Considering the gains acquired as a result of mining, it has become a considerably sensitive issue for companies to protect and process the data they have and received from the outside.

Data mining algorithms have different default parameters which can be changeable to give the convenient result. Hyperparameter optimization is important for making enhancement on data mining algorithms. In this study, we aim to improve performance of algorithms with optimizing hyperparameters.

Car evaluation dataset has been used by applying different data mining algorithms in this study. Many predictive models have been developed with the same purpose of our study using this dataset (detailed information on this dataset is available in Section 2.2) in literature. Makki and Mustapha use Backpropagation Neural Network and Naïve Bayesian algorithms with different test mode on this dataset for comparison. Neural Network model with 10 fold cross validation has highest score (99.53%) (Makki 2011). Awwalu applies three different algorithms, namely ANN, Naïve Bayes and Decision Tree. In addition, the study compares speed and accuracy of the algorithms. ANN and Naïve Bayes give 93.51% accuracy with 10-fold cross validation technique (Awwalu,2014). Frayman implements a fuzzy neural network consisting of four layers, which are: the input, input membership function, rule and output layers. Frayman's study compares FNN and C4.5 results indicating that FNN (92.7%) has provided better accuracy rate when compared to C4.5 (88.4%) (Frayman,1999). Moutinho's study makes an analysis on data related to car purchase in order to show impact of gender on car buyer [4]. There are six attributes are used in the study applying a neural network. Moutinho's

study provides the importance level of attributes with reasons according to gender car prices for different reasons. Female persons thought that if they bought expensive car it would perform as it should. Expectations and impression are reasons for selecting according to price of a car for male persons. Also there are more studies related to car purchasing (Moutinho,1996;Jørgensen,1990;Jansson,1989). Algorithm performances that are used on different car datasets can be improved with feature selection techniques and tuning hyperparameters. In literature, current studies (using same car evaluation dataset) do not apply hyperparameter optimization on algorithms. Thus, this study aims to show improvement performances with tuning hyperparameters.

Hyperparameter optimization is the selection of most suitable parameter values from parameter space. In machine learning, each algorithm has its own parameters. Correct parameters cannot be learned directly from these algorithms, so it provides us to optimize model complexity. Selecting correct parameters improves performance of algorithm. Chris Tornton (Thornton,2013) used 21 popular datasets from the UCI repository to compare results of classifiers with different types that applied feature selection and parameter optimization techniques. They highlight that performance of classifiers with default values is mostly much better than using parameter optimization or feature selection techniques. Jan N. van Rijn (van Rijn,2018) tried to find most important hyperparameters on different datasets for three different algorithms which are random forest, support vector machines and adaboost. According to prediction results, the gamma and complexity parameters for support vector machines, the maximum depth and learning rate parameters for adaboost, the minimum number of samples per leaf and maximum features available for a split parameters for random forest are chosen as the most important parameters. Rayrone Z. N. Marques (Marques,2015) apply hyperparameter optimization on 28 datasets for different 6 algorithms which are J48, RandomTree, RepTree, JRip, PART and Ridor. They used EMiner tool (Elder,1998) for hyperparameter optimization. Their study reveals that when this tool is used in order to hyperparameter selection, the study has provided better results on algorithm performances. The results can be improved by using different optimization techniques such as Bayesian optimization (Snoek,2012).

The rest of the paper is organized as follows: Section 2 presents the method and material included prediction phases, classification algorithms. Section 3 shows results of our study. Finally, conclusions and comment on future work are presented in section 4.

2. METHODS AND MATERIALS

2.1 Workflow of the Process

Workflow of the process can be seen in Figure 1. Firstly, we obtain our dataset from the University of California Irvine (UCI). Dataset is separated as train and validation set with five fold. Then classifiers which are described in Section 2.3 are selected. For each classifier their own parameters are tuned and values that give best results are taken. Classifiers with their best hyperparameters are modeled and performed on dataset. Finally, prediction results are compared.

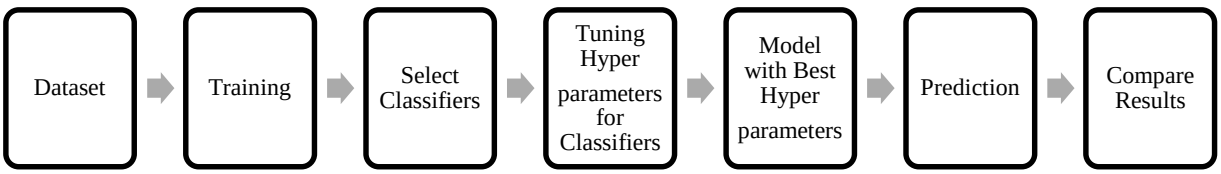

Figure1. Prediction Phase for Classification

2.2 Dataset and Training

We train algorithms using the car evaluation dataset. The car dataset is obtained from the University of California Irvine (UCI) as specified in Section 2.1. This dataset contains six attributes and four classes, totally has 1728 instances. The attributes are buying price, maintenance, number of doors, number of persons, luggage capacity and safety. The values for each attribute are different and mostly nominal. Nominal values are converted to numeric values. In order to compare the results of the algorithms, 5 fold cross validation was performed. We take different fold number between 4 and 10 to divide dataset. The number of five has given the best accuracy so dataset is divided into five equal sizes. One of these pieces is taken as test set and others are taken as train set.

2.3. Methods

Algorithms applied to our dataset are chosen according to similar datasets in literature. The algorithms are support vector machines, k-nearest neighbor, gradient boosting, random forest, multilayer perceptron have performed.

2.3.1 Support Vector Machines

Support Vector Machines is a simple and effective method used in classification. Its main purpose is to simplify the data and provide more understandable information to the user. Support Vector Machines are basically used to separate the data of the two classes in the most appropriate way. In this case, decision boundaries or in other words, hyper planes are determined (Vapnik,1995). It is separated as linear and nonlinear support vector machines. In a nonlinear dataset, DVMs cannot draw a linear hyper-plane. For this reason, kernel tricks called kernel numbers are used. The kernel method greatly increases machine learning in nonlinear data.

2.3.2 K-Nearest Neighbor

The main principle of k nearest neighbor algorithm is based on finding the samples closest to the sample to be classified. The classification of the data we have is calculated according to the threshold value that occurs with the average of the k data determined to be the closest by calculating the common properties of the data in the learning set (Peterson, 2009). As we know the class of the data we have in advance, it is supervised learning.

2.3.3 Gradient Boosting

Gradient Boosting is a machine learning technique for regression and classification problems. This algorithm creates a model formed by decision trees by combining weak prediction models. The main purpose of the algorithm is to define and minimize a loss function. It iteratively solves residuals. Gradient Boosting creates a function (G) that generates predictions at the first iteration. It calculates the difference between estimates and target value and creates a different function (K) for these differences. In the second iteration, it combines these functions (G and K) and calculates the difference between estimates and targets again. In this way, it constantly tries to increase the success of the "G" function by adding it, thereby reducing the difference between predictions and targets to zero. By this way, it updates the parameters of the learning model (Friedman, 2002).

2.3.4 Random Forest

Random forest algorithm is a type of ensemble learning algorithms. The random forest algorithm is an algorithm that aims to increase the classification value by generating multiple decision trees during

the classification process. The individually created decision trees form a decision forest. The decision trees here are a subset of randomly selected data sets. Classification decision trees established differently from each other constitute the decision forest community that will lead us to the result. The results obtained during the decision forest formation are combined and the last estimate is made (Liaw,2002).

2.3.5 Multilayer Perceptron

Multilayer perceptron is a neural network model. MLP consist of one input layer, one or more hidden layer and one output layer. The information in the input layer is transmitted to the output of the network by processing each cell individually. It has very common usage such as classification and character recognition (Balakrishnan,2005;Orhan,2011;Verma,1995;Patil,2011; Sabourin,1992).

2.4 Hyperparameter Optimization and Proposed Model

The parameters of each algorithm (as specified in Section 2.3) have been optimized in our study. The C and gamma parameters of the SVM (Duan,2003), number of neighbors and distance metric parameters of the k-nearest neighbor, number of trees parameter of random forest (Bernard,2009), the number of maximum depth and estimator parameters of the gradient boosting (Natekin,2013), the hidden layer size and solver parameters of multilayer perceptron are optimized.

C and Gamma are two important hyper parameters of support vector machines algorithm. C is the cost parameter that gives cost of misclassification. C parameter with high value gives low bias and high variance. Because of penalizing cost of misclassification a lot, low bias occurs. C parameter with low values gives high bias and low variance. Gamma is a parameter of Gaussian kernel. Gamma parameter with low value means a Gaussian with a large variance and low bias, while gamma parameter with high value means a Gaussian with a low variance and high bias (Ben-Hur,2010).

Number of neighbors and distance metric are parameters of KNN algorithm. Number of neighbors defines closest number of neighbors according to distance. Distance metric calculates distance between data. Three type of distance metrics are generally used for calculation, namely Euclidean, Manhattan and Minkowski (Chomboon,2015).

Number of estimators and maximum depth are parameters for gradient boosting algorithm. Number of estimators defines number of boosting stages or number of trees in to be modeled. Maximum depth defines largest depth of a tree. The high number of depth makes more splits on tree, so it captures more information about how the data. In addition, number of estimator parameter defines number of trees in the random forest algorithm.

The hidden layer size and solver are parameters for MLP. The size of the hidden layer is the size of nodes between input and output layers in multilayer perceptron algorithm. The solver parameter is used for weight optimization. There are different types of solver such as Adaptive Moment Estimation (Kingma,2014), Stochastic Gradient Descent (Bottou,2010).

Different values are given for each parameter. The values that give best results are selected and then applied to algorithm. The purpose of the hyperparameter optimization is to show that algorithms can achieve better results with optimized values, not their default values.

3. RESULTS

Table 1 summarizes importance of each attributes in dataset. Safety is most important feature in car evaluation. It is followed by number of persons.

Table1. The Importance of Attributes

Attributes	Importance
Buying Price	0.16469782
Maintenance	0.16226334
Number of Doors	0.05786422
Number of Persons	0.25233984
Luggage Capacity	0.08045919
Safety	0.28237559

Table 2 shows the accuracies of all algorithms with default parameters (without optimization) and with hyperparameter optimization. According to default parameter results, gradient boosting has best prediction result. It followed by SVM. Gradient boosting with hyper parameter optimization method produces best predictions of the car evaluation by 99.42%. KNN gives best improvement as a comparison with and without hyper-parameter optimization.

Table2. Classification Accuracy of All Algorithms with and without Hyper-parameter Optimization

Model	Prediction Accuracy Without Optimization	Prediction Accuracy With Optimization
SVM	97.00	98.9
RandomForest	96.80	97.9
KNN	92.40	95.4
Gradient Boosting	99.30	99.4
MLP	92.25	93.0

According to these results, safety takes first place for deciding buying a car. Also, number of persons is critical for evaluation. On the other hand, parameter optimization seems to improve the results. Although the gradient boosting algorithm gives the best results, it is seen that the k-nearest neighbor algorithm gives better results when the amount of improvement is considered.

In terms of the results, safety is the most crucial feature. Automotive safety aims to minimize the occurrence and consequences of accidents involving motor vehicles. It is expected to be the most important attribute, since it is related to life safety. Nevertheless, considering less developed country the importance of features can be changeable. In countries with low economic power, it is expected that the price will become a more important feature when people are buying vehicles (Khan,2016).

Our study can be seen as a guide for people in buying and selling cars. Car dealers have information about acceptability. They can procure vehicles in line with acceptance and increase sales. With adding more relative attributes like price and amount of petrol used (Bendtsen,1980), more extensive studies can be done.

4. CONCLUSION

In this study, different algorithms are applied on the car evaluation dataset. Hyper-parameter optimization was performed on support vector machines, k-nearest neighbor, random forest, multilayer perceptron and gradient boosting algorithms. The calculations highlight that gradient boosting with hyper parameter optimization method produces best predictions of the car evaluation by 99.42%. More clear results can be obtained with an extended data set with different parameters.

REFERENCES

Awwalu, J., Ghazvini, A., & Bakar, A. A. (2014). Performance Comparison of Data Mining Algorithms: A Case Study on Car Evaluation Dataset. Int. Jour. of Computer Trends and Technology (IJCTT), 13(2).

Balakrishnan, D., & Puthusserypady, S. (2005, April). Multilayer perceptrons for the classification of brain computer interface data. In Proceedings of the IEEE 31st Annual Northeast Bioengineering Conference, 2005. (pp. 118-119). IEEE.

Bendtsen, P. H. (1980). The influence of price of petrol and of cars on the amount of automobile traffic. International Journal of Transport Economics/Rivista internazionale di economia dei trasporti, 207-213.

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. In Data mining techniques for the life sciences (pp. 223-239). Humana Press.

Bernard, S., Heutte, L., & Adam, S. (2009, June). Influence of hyperparameters on random forest accuracy. In International Workshop on Multiple Classifier Systems (pp. 171-180). Springer, Berlin, Heidelberg.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010 (pp. 177-186). Physica-Verlag HD.

Chomboon, K., Chujai, P., Teerarassamee, P., Kerdprasop, K., & Kerdprasop, N. (2015, March). An empirical study of distance metrics for k-nearest neighbor algorithm. In Proceedings of the 3rd international conference on industrial application engineering (pp. 1-6).

Duan, K., Keerthi, S. S., & Poo, A. N. (2003). Evaluation of simple performance measures for tuning SVM hyperparameters. Neurocomputing, 51, 41-59.

Elder, John F., and Dean W. Abbott. "A comparison of leading data mining tools." Fourth International Conference on Knowledge Discovery and Data Mining. Vol. 28. 1998.

Frayman, Y., Ting, K. M., & Wang, L. (1999, July). A fuzzy neural network for data mining: dealing with the problem of small disjuncts. In IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339) (Vol. 4, pp. 2490-2493). IEEE.

Friedman, J. H. (2002). Stochastic gradient boosting. Computational statistics & data analysis, 38(4), 367-378.

Jansson, Jan Owen. "Car demand modelling and forecasting: a new approach." Journal of transport Economics and Policy (1989): 125-140.

Jørgensen, Finn, and Tore Wentzel-Larsen. "Forecasting car holding, scrappage and new car purchase in Norway." Journal of Transport Economics and Policy (1990): 139-156.

Khan, M., & Sinclair, M. (2016). Importance of safety features to new car buyers in South Africa. Southern African Transport Conference.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lapersonne, Eric, Gilles Laurent, and Jean-Jacques Le Goff. "Consideration sets of size one: An empirical investigation of automobile purchases." International Journal of research in Marketing 12.1 (1995): 55-66.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.

Makki, S., Mustapha, A., Kassim, J. M., Gharayebeh, E. H., & Alhazmi, M. (2011, April). Employing neural network and naive Bayesian classifier in mining data for car evaluation. In Proc. ICGST AIML-11 Conference (pp. 113-119).

Marques, Rayrone ZN, et al. "An experimental evaluation of data mining algorithms using hyperparameter optimization." 2015 Fourteenth Mexican International Conference on Artificial Intelligence (MICAI). IEEE, 2015.

Moutinho, Luiz, Fiona Davies, and Bruce Curry. "The impact of gender on car buyer satisfaction and loyalty: A neural network analysis." Journal of Retailing and Consumer Services 3.3 (1996): 135-144.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, 21.

Orhan, U., Hekim, M., & Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. Expert Systems with Applications, 38(10), 13475-13481.

Patil, V., & Shimpi, S. (2011). Handwritten English character recognition using neural network. Elixir Comput Sci Eng, 41, 5587-5591.

Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883.

Sabourin, M., & Mitiche, A. (1992). Optical character recognition by a neural network. Neural networks, 5(5), 843-852.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In Advances in neural information processing systems (pp. 2951-2959).

Thornton, Chris, et al. "Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013.

van Rijn, Jan N., and Frank Hutter. "Hyperparameter importance across datasets." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

Vapnik, V.N., 1995, The Nature of Statistical Learning Theory, Springer-Verlag, New York.

Verma, B. K. (1995, November). Handwritten Hindi character recognition using multilayer perceptron and radial basis function neural networks. In Proceedings of ICNN'95-International Conference on Neural Networks (Vol. 4, pp. 2111-2115). IEEE.