International Journal of Computer Applications Technology and Research
Volume 3– Issue 6, 349 - 352, 2014, ISSN: 2319–8656

Proposing an Appropriate Pattern for Car Detection by Using Intelligent Algorithms

Amin Ashir
The member of young researchers club, Islamic Azad University of Dezful, Iran

Sedigheh Navaezadeh
Sama Technical and Vocational Training College, Islamic Azad University, Mahshahr, Branch Mahshahr, Iran

Abolfazl jafari
Sama Technical and Vocational Training College, Islamic Azad University, Mahshahr, Branch Mahshahr, Iran

Abstract: Nowadays, the automotive industry has attracted the attention of consumers, and product quality is considered as an essential element in current competitive markets. Security and comfort are the main criteria and parameters of selecting a car. Therefore, standard dataset of CAR involving six features and characteristics and 1728 instances have been used. In this paper, it has been tried to select a car with the best characteristics by using intelligent algorithms (Random Forest, J48, SVM, NaiveBayse) and combining these algorithms with aggregated classifiers such as Bagging and AdaBoostMI. In this study, speed and accuracy of intelligent algorithms in identifying the best car have been taken into account.

Keywords: intelligent algorithms, pattern recognition, Random Forest, J48, SVM, Naïve Bayse, AdaBoostMI

1. INTRODUCTION

Nowadays, the automotive industry has been considered by consumers, and the product quality has been recognized as an essential element in the current competitive markets. Nowadays, in business process, consumer has a crucial role in improving and developing the product, and is considered as a base for receiving reliable information. Security and comfort are important criteria and parameters in selecting a car. The framework of this paper is on the basis of quality management. Also, in this study, those indicators that have considered the consumer have been taken into account. Therefore, standard database of CED [1] has been used. In this database, there are various criteria such as security and comfort. In the rest of this paper, the criteria of car identification, investigation of intelligent algorithms, layout and the results of simulation have been respectively analyzed.

2. CRITERIA OF IDENTIFYING THE CAR IN TERMS OF SECURITY AND COMFORT

The criteria that have been used to identify the cars are as follows:

Car prices;

Maintenance cost;

The number of car doors;

The car capacity (the number of passengers);

The required space for furniture;

Security of the car.

3. AGGREGATED CLASSIFIERS AND METHODS

Generally, searching is carried out in the imaginary space in supervised learning algorithms so that an appropriate prediction can be considered to find a solution. An aggregated classifier is a supervised learning algorithm combining various theories so that a better theory is presented. As a result, an aggregated classifier is considered as a technique combining weak learners so that a strong learner is created. Fast algorithms such as decision trees are considered along with aggregated classifiers. Observations show that when diversity of models is great, aggregated classifiers perform efficiently. Therefore, different methods have been proposed to increase the diversity among combined models. The most well-known methods are Bagging and Boosting [2]. In Bagging method, the classifiers designed in various versions of data are combined, and majority voting is individually considered among classification decisions. Since re-sampling of Bootstrap is usually used to locate initial data due to imbalanced data, this method is called Bootstrap Aggregation or Bagging. One of the classifiers usi9ng Bagging and AdaBoostMI methods is Random Forest involving several decision trees, and its output is obtained through individual output trees. This algorithm combines the Bagging method by random selection of features so that a set of decision trees with controlled diversity are created. One of the advantages is high accuracy of the classifier. Also, it can perform well with many outputs [3]. The second wee-known method is Boosting that trains new samples and instances to reinforce learning samples and instances, and it makes some changes in aggregated classifier. This method has sometimes higher precision and accuracy in comparison to Bagging method. One of its disadvantages is maximum learning and training of these learning samples and instances. AdaBoostMI is one of the well-known Boosting methods.

4. SVM CLASSIFIER

This algorithm is a supervised learning method for classification and regression. This method is relatively a new method. In recent years, it has better efficiency that older method in terms of classification such as perceptron neural networks. The framework of SVM classifier is on the basis of linear classification of data. In learner classification of data, it has been tried to select a line that has more reliability margin. A solution can be found for optimization line of data through QP methods that are well-known methods in solving limited problems. Before dividing a line, data should be transferred to a space with higher dimensions so that machine can classify very complex data [4, 5].

This algorithm has theoretical foundations. It just requires a dozen of samples, and is not sensitive to the number of problem dimensions. In addition, efficient methods of SVM learning have been considerably developed. In learning process involving two classes, the aim of SVM is to find the best function for classification so that the members of two classes can be determined and distinguished in dataset. The criteria of the best classification are determined geometrically. This is true for data set analyzed linearly. The boundary defined as a part of space or separation of two classes can be defined by hyperplane. This geometric definition allows us to detect that how the boundaries are maximized, even if there are numerous hyperplanes, and only a few of them are considered as a solution of SVM. SVM persists on the largest boundary for hyperplane since it provides generalization of algorithm as well as possible. This issue helps to classification efficiency and accuracy in tested data. Also, it provides a space for better classification of future data. One of the problems of SVM is that it requires complex computations. However, this problem has been acceptably solved. One solution is that a large optimization problem can be divided into smaller problems. Each problem involves a pair precisely selected from variables. These variables can be used in problems. This process continues until all analyzed sections are solved.

5. CLASSIFIER OF NAIVE BAYES

Naive Bayes presents a predictive model in terms of output probability/special results. NB algorithms measure the patterns or the relation between data by counting the number of observations. Then, this algorithm provides a model demonstrating the patterns and their relations. After creating this model, it can be used as a predictive model. This algorithm helps us to present models for classification and prediction of various objectives. Notice the following example:

Which customers are interested in buying a special product.

Which consumers can purchase more than 10000 dollars.

Identifying the customers who do not buy the company products, and buy the products of competitors.

Predicting products that have failed, and predicting their probability.

Naïve Bayes algorithm predicts the above mentioned issues by Bayes theory supposing that data values are independent. This algorithm provides a model as quickly as possible, and it can be considered for classifying two or more than two classes [6].

6. J48 CLASSIFIER

Algorithm Description

J48 is not just an algorithm; rather, it is a set of algorithms. The performance of J48 has been described and demonstrated in figure 1. All methods of tree deduction begin from root node providing all data information, and recursively divides information into smaller sections. In this case, each ratio is tested in each node. Sub-trees show the classification of main information completing tests of determined ratio valuation. This process continues until all sets are purified; that is, all samples are placed in a group, and the growth of tree stops in this time.

The rules of classifiers

J48 introduces a list of rules in the frame of a form. Rules are designed for each class. A sample returns to its secure location by finding the primary rules. If a rule cannot be found for a sample, then it is located in presupposition class. In the rules of J48, time of CPU and required memory are considered.

Input: an attribute-valued dataset D

Tree={}

If D is "pure" OR other stopping criteria met then

Terminate

End if

For all attribute aϵ D do

Compute information-theoretic criteria if we split on a

End for

abest = Best attribute according to above computed criteria

Tree = Create a decision node that tests abest in the root

Ds = Induced sub-datasets from D based on abest

For all Dsdo

Treec = J48(Dv)

Attach Treec to the corresponding branch of tree

End for

Return Tree

Figure 1: the performance of J48 algorithm

7. LAYOUT

Data used in this paper are a set of data related to car entitled CED [1]. Collected data have six characteristics and 1728 instances. The method of this study is on the basis of pattern recognition, and it has been tried to detect and identify the best car by using intelligent algorithms. It should be mentioned that intelligent algorithms have been executed in Weka software [7].

The used instances are divided into 10 parts. By increasing this division, better results can be obtained, but it is time consuming. In the rest of this paper, the obtained results will be analyzed and studied.

8. SIMULATION RESULTS

After applying intelligent algorithms on the instances, the obtained results have been collected in three tables. Table 1 demonstrates the comparison of intelligent algorithms (Random Forest, J48, SVM, NaiveBayse) in terms of time and accuracy in identifying a secure and comfortable car.

In table 2, these algorithms have been compared with aggregated classifier of Bagging. Also in table 3, these algorithms have been compared with aggregated classifier of AdaBoostMI to identify the best car, and the criteria of time and accuracy have been investigated.

Table 1: comparing intelligent algorithms in terms of time and accuracy in secure and comfortable cars

Accuracy	**Time**	Criteria / algorithms
%93.75	1.81 Sec	**SVM**
%85.5324	0.02 Sec	**Naive Bayse**
%92.3611	0.06 Sec	**J48**
%96.4606	0.13 Sec	**Random Forest**

As it is observed in table 1, the accuracy of time algorithm and speed of Naïve Bayes algorithm are better than other algorithms.

Table 2: comparing intelligent algorithms with aggregated classifier of Bagging in terms of time and accuracy in identifying cars.

Accuracy	**Time**	Criteria / algorithms
%93.3449	13.16 Sec	**SVM**
%85.3009	0.09 Sec	**Naive Bayse**
%93.8079	0.02 Sec	**J48**
%45.5023	0.92 Sec	**Random Forest**

As it is observed in table 2, the accuracy of used algorithms has not been improved by using aggregated classifier of Bagging. In most of these algorithms, accuracy has been reduced after using aggregated classifier of AdaBoostMI.

Table 3: comparing intelligent algorithms with aggregated classifier of AdaBoostMI in terms of time and accuracy in identifying secure and comfortable cars.

Accuracy	**Time**	Criteria / algorithms
%94.5023	21.80 Sec	**SVM**
%90.162	0.33 Sec	**Naive Bayse**
%961227	0.45 Sec	**J48**
%93.6921	0.14 Sec	**Random Forest**

As it is observed in table 3, accuracy of used algorithms has been improved by using aggregated classifier of AdaBoostMI.

Accuracy of J48 algorithm is above 96%. Comparing tables 2, 1 and 3 show that, by using aggregated classifier of AdaBoost, accuracy of J48 algorithm is above 96%, and 0.45 seconds has the best accuracy. In addition, in naïve Bayes algorithm without aggregated classifier of bagging and AdaBoostMI, 0.02 seconds is considered as the best time in identifying secure and comfortable cars. In J48 algorithm, 0.02 seconds is considered as the best time.

9. CONCLUSION

In order to identify a secure and comfortable car, considerable results can be obtained by considering standard CEO database and intelligent algorithms such as SVM, Naïve Bayes, J48, Random Forest as well as combining these algorithms with aggregated classifiers of Bagging and AdaBoostMI. In this paper, since evaluating simulation results on the basis of identification time and modeling accuracy and accuracy in identifying secure and comfortable car have great importance, it is better to use J48 algorithm with aggregated classifier of AdaBoostMI involving accuracy of 96.1227 percent and time of 0.45 seconds.

11. REFERENCES

[1] http://archive.ics.uci.edu/ml/datasets/Car+Evaluation, December 2011

[2] DietterichThomas G. "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization", Machine Learning 40(2), 139–158, 2000.

[3] KhoshgoftaarT.M., GolawalaM. Van HulseJ., "An Empirical Study of Learning from Imbalanced Data Using

Random Forest", Proceedings of the 19th.IEEE Conference on Tools with Artificial Intelligence, pp.310-317, 2007.

[4] Vapnik V., "Statistical Learning Theory", John Wiley & Sons, New York, 1998.

[5] Vapnik V., "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 1995.

[6] Jam Sahar, Khashayar, 1390. An approach of smart business, data warehouse and data mining.www.hamsahar.com

[7] http://www.cs.waikato.ac.nz/~ml/weka, November 2011.