International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-11, September 2020

Performance Enhancement Method for Machine Learning Algorithm

Archana Chaudhary

Abstract*: Machine learning is programming computer or a mobile device that learns from experience. Machine learning classification methods are helpful in various fields of Computer Science like driverless cars, product recommendation systems, dynamic pricing, Google translate, online video streaming, internet and mobile fraud detection systems and much more. The present work proposes a method augClassifier to enhance the performance of Simple Logistics machine learning algorithm. The performance assessment of machine learning algorithm is conducted on a Mobile device using Android Environment. The work also presents the comparative performance investigations of Simple Logistics machine learning algorithm using correlation based feature selection method with respect to performance measures Precision, Sensitivity, F-Measure and ROC. The present work conforms that the augClassifier enhances the performance of Simple Logistics machine learning algorithm.

Keywords*: Classification, *Machine learning, Precision, Sensitivity.

I. INTRODUCTION

Machine learning is a modern research area of Artificial Intelligence, in which computers are instructed to copy human intellect. In this field a computer or a mobile device is programmed to improve the performance using some example data or past experience [1]. It exercises statistical principles for developing models. The prime objective of any machine learning model is to infer or reason from past experience [12, 17, 18, 19, 20, 21, 22]. Machine learning applications are widely used in different fields [6, 7, 8]. Individual human travel patterns were quickly traced by a mobile phone in [2]. Function points of software were assessed using Support Vector Machine (SVM), Artificial Neural Network (ANN) machine learning algorithms [3]. The Experiments performed in [3] showed that ANN and SVM confirmed as efficient methods for predicting the function points. Machine learning algorithms SVM, Naive Bayes and Adaboost were applied on Facebook dataset for identification of cybercrime [4]. SVM proved to be an outstanding algorithm as compared to the other algorithms [4]. The ability of a mobile learner of English was improved by a mobile intelligent system [5]. The intelligent system approved English news articles to the learners on the basis of learner's capabilities.

Revised Manuscript Received on September 30, 2020.
***** Correspondence Author
Archana Chaudhary*, School of Computer Science & IT, Devi Ahilya University, Khandwa Road, Indore 452001, Madhya Pradesh, India. Email: archana_scs@yahoo.in

II. MATERIALS AND METHODS

The main focus of feature selection methods is to search feature subsets that would scale up the classification accuracy obtained from machine learning algorithm. The feature selection methods remove the irrelevant features from feature set [10, 15]. The feature selection method used in the present work is as under –

A. Correlation based feature selection method

In the present work correlation based feature selection method is used. The method exercises a guiding function for extracting feature subsets [9]. The guiding function is also known as subset evaluation function. It isexpressed as under

$$Grade_s = \frac{kn \times \overline{r_{feature-class}}}{\sqrt{kn+(kn-1)\times\overline{r_{f-f}}}} \qquad (1)$$

where $k_n$represents the number of features in a given subset s, the average feature–class correlation is represented by $r_{feature\text{-}class}$, the mean feature–feature inter-correlation is represented by $r_{f\text{-}f}$ and $Grade_s$symbolizes the worth of S.

B. Performance measures

The performance analysis of any Machine learning algorithm can be conducted with the help of performance measures [11]. In order to evaluate the performance of machine learning algorithm the performance measures –Classification accuracy, Precision, Sensitivity, F-Measure and Receiver Operating Characteristics (ROC) are considered in the present work.

C. Machine learning Algorithm

An instance represents an example and has a set of characteristic features.A machine learning classification problem allocates a class or a category to an instance with unknown class [9]. A machine learning classification algorithm searches for a class of an unlabelled instance with the help of examples of a training set.

- **Simple Logistics (SLO)**

It is a regression method that is used for classification in machine learning [16]. It is a state of the art machine learning algorithm. The algorithm assumes classes as binary classes for classification. The probability that an example belongs to a particular class using SLO is given by prob1(x), as shown in Eqn. (2) for abinary type of classification problem. In Eqn. (2) shown below a and bsignify the algorithm parameters –

$$prob1(x) = \frac{e^{a+b'\ x}}{1+e^{a+b\ \ x}} \qquad (2)$$

- **The training dataset**

The dataset in [13] is enriched by adding symptoms of 4 more crop diseases like Rhizoctonia foliar blight, Anthracnose, Scab and Verticillium wilt. After appending the aforementioned diseases the crop disease dataset consists of 17 disease classes and 2022 instances. The dataset has no missing feature values.

III. THE PROPOSED METHOD

In this work a new method – *augClassifier* for enhancing the performance of SLO machine learning algorithmis suggested. The method is shown in Algorithm 1 below -

Algorithm 1.*augClassifier*.

Input: $D_{Training}$ = {e_1, e_2, . . . ,e_n} // Training set which
// consists of training examples and their class labels.

Output: Classification accuracy.

Step1. Apply correlation based feature selection method on $D_{Training}$ and extract importantfeatures.

Step2. Perform arbitrary sampling on the result obtained from (Step 1) and attain $D_{cfs\text{-}resample}$.

Step3. Apply a machine learning classification algorithm.
// SLO machine learning algorithmis applied.

Step4. Calculate classification accuracy from the Step 3.

Step 5. Output classification accuracy.

The input to the proposed augClassifier is a training set $D_{Training}$ which contains the training examples of crop diseases.

The important features are extracted using correlation based feature selection method using Step 1 of Algorithm 1. In Step 2 of Algorithm1, arbitrary sampling is done and a uniform dataset $D_{cfs\text{-}resample}$ is generated.

In Step 3 of Algorithm 1, the SLO machine learning algorithm is applied on the dataset $D_{cfs\text{-}resample}$acquired from the earlier step.

The resultant classification accuracy is obtainedfrom Step 5 of Algorithm 1.

IV. RESULTS AND DISCUSSION

Experiments are conducted on a mobile device with the specifications of [14]. The present work is conducted with the help of 10 folds cross validation strategy in which the whole crop disease dataset is partitioned into 10 equal sized subsets and classification algorithm is trained on 9 subsets and then tested on the left over subset. The disease classification accuracy observations are shown in Table-I. It is evident from Table-I that method *augClassifier* enhances the disease classification accuracy of SLO machine learning algorithm.

Table- I: Disease classification accuracies using machine learning algorithms

Machine learning algorithm	Disease classification accuracy (in %)
SLO	93
augClassifier	96

It is also clear from Fig.1 that the *augClassifier* outperforms the original SLO machine learning classification algorithm.

Further it is evident from Fig.1that for performance measures - Precision, Sensitivity, F-Measure and ROC, the *augClassifier* outperforms the original SLO machine learning algorithm.

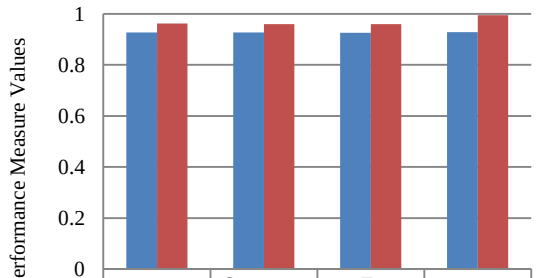

Fig.1.Performance comparison of AugClassifier and SLO.

V. CONCLUSIONS

The present work proposed a new method - *augClassifier* to enhance the performance of SLO machine learning algorithm. The performance assessment was conducted on a mobile device using Android Environment. The *augClassifier* method resulted in disease classification accuracy as 96%, whereas the original SLO showed classification accuracy as 93%. The *augClassifier*enhanced the disease classification accuracy of SLO machine learning algorithm. The *augClassifier* method also showed performance enhancement as compared to SLO with respect to performance measures Precision, Sensitivity, F-Measure and ROC. Hence the *augClassifier* can be used with SLO for disease diagnostic applications.

REFERENCES

1. E. Alpaydin, “Introduction to Machine Learning”, Second Edition, MIT press, Cambridge, London, (2009).
2. F. Liu, D. Janssens, G. Wets, M. Cools, “Annotating mobile phone location data with activity purposes using machine learning algorithms”, Expert Systems with Applications, Volume 40, (2013), pp. 3299–3311.
3. G.Sikka, A. Kaur, M. Uddin, “Estimating Function points: Using Machine Learning and Regression Models”, 2nd International Conference on Education Technology and Computer (ICETC), Shanghai, (2010).
4. Y.P.Singh, H.M. Deylami, “Cybercrime detection techniques based on Support Vector Machines”, Artificial Intelligence Research, vol. 2, no. 1, (2013).
5. C. M. Chen, S. H. Hsu, “Personalized Intelligent Mobile Learning System for Supporting Effective English Learning”, Educational Technology & Society, vol. 11, no. 3, (2008), pp. 153-180.
6. R.Eisinger, R. Romero, R. Goularte, “Machine Learning Techniques Applied to Dynamic Video Adapting”, IEEE Seventh International conference on Machine Learning and Applications(ICMLA '08), (2008).
7. A.Forster, “Machine Learning Techniques Applied to Wireless Ad-Hoc Networks: Guide and Survey”, International conference on Intelligent Sensors, Sensor Networks and Information, Melbourne, Queensland, (2007), pp. 365 – 370.
8. M. Slavik, I. Mahgoub, “Applying Machine Learning to the Design of Multi-Hop Broadcast Protocols for VANET”, Seventh International Conference on Wireless Communications and Mobile Computing (IWCMC), (2011), pp. 1742 – 1747.
9. L. Silva, M. Koga, C. Cugnasca, A. Costa, “Comparative assessment of feature selection and classification techniques for visual inspection of pot plant seedling”, Computers and Electronics in Agriculture, vol. 97, (2013), pp. 47–55.

10. A.Sharma, S. Dey, "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis features", Special Issue of International Journal of Computer Applications (0975 – 8887) on Advanced Computing and Communication Technologies for HPC Applications, (2012).
11. M.Sokolova, G. Lapalme, "A systematic analysis of performance measures for classification tasks", Information processing and Management, vol. 45, (2009), pp. 427-437.
12. A.P. Tonda, E. Lutton, R. Reuillon, G. Squillero, P.H. Wuillemin, "Bayesian Network Structure Learning from Limited Datasets through Graph Evolution", Proceedings of 15th European Conference on Genetic Programming (EuroGP), vol. 7244, (2012), pp. 254-265.
13. A. Chaudhary, S. Kolhe, R. Kamal, "An improved random forest classifier for multi-class classification", Information Processing in Agriculture, (2016), pp. 215 – 222.
14. P.Liu, Y. Chen, W. Tang, Q. Yue, "Mobile WEKA as Data Mining Tool on Android", Advances in Intelligent and Soft Computing, vol.139, (2012), pp. 75- 80.
15. H.Liu, H. Motoda, "Computational methods of feature selection", Chapman and Hall/CRC, Taylor and Francis Group, NW, (2008).
16. T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning: Prediction, Inference and Data Mining", 2nd edition, Springer-Verlag, New York, (2009).
17. A. Chaudhary, S. Kolhe, R. Kamal, "A hybrid ensemble for classification in multiclass datasets: An application to oilseed disease dataset", Computers and Electronics in Agriculture, vol. 124, (2016), pp. 65–72.
18. A. Chaudhary, S. Kolhe, R. Kamal, "Machine learning techniques for mobile intelligent systems: A study", In IEEE Ninth International Conference on Wireless and Optical Communications Networks (WOCN), (2012), pp. 1–55.
19. A. Chaudhary, S. Kolhe, R. Kamal, "Machine Learning Classification Techniques: A Comparative Study", International Journal on Advanced Computer Theory and Engineering, vol. 2, no. 4, (2013), pp. 21–25.
20. A. Chaudhary, S. Kolhe, R. Kamal, "Machine Learning Techniques for Mobile Devices: A Review", International journal of Engineering Research and Applications, vol. 3, no. 6, (2013), pp. 913–917.
21. A. Chaudhary, S. Kolhe, R. Kamal, "Performance Examination of Feature Selection methods with Machine learning classifiers on mobile devices", International Journal of Engineering Research and Applications, vol. 3, no. 6, (2013), pp. 587–594.
22. A. Thakur, R. Thakur, "Machine Learning Algorithms for Intelligent Mobile Systems", International Journal of Computer Sciences and Engineering, vol. 6, no. 6, (2018), pp. 1257–1261.

AUTHOR'S PROFILE

Archana Chaudhary (Thakur)obtained her Ph.D. in Computer Science from Devi Ahilya University, Indore, India in 2016. She is serving as an Assistant Professor at School of Computer Science & IT, Devi Ahilya University, Indore. She is supervising many doctoral scholars. She has published various research works in different national and international journals. She has also authored various original scientific research articles in Elsevier journals. She has also served as areviewer of Elsevier journal manuscripts. She has actively participated and presented research papers in different international and national conferences. She has also been esteemed speaker in IEEE international conferences. She has also chaired sessions in IEEE international conferences. Her research areas include Artificial Intelligence, Machine learning, Data Mining and Soft Computing.