Computers and Electronics in Agriculture 170 (2020) 105247

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A particle swarm optimization based ensemble for vegetable crop disease recognition

Archana Chaudhary[a,*], Ramesh Thakur[b], Savita Kolhe[c], Raj Kamal[d]

[a] School of Computer Science & IT, Devi Ahilya University, Khandwa Road, Indore 452001, Madhya Pradesh, India
[b] International Institute of Professional Studies, Devi Ahilya University, Khandwa Road, Indore 452001, Madhya Pradesh, India
[c] ICAR – Indian Institute of Soybean Research, Department of Agricultural Research and Education, Khandwa Road, Indore 452 017, Madhya Pradesh, India
[d] Prestige Institute of Engineering Management and Research, Indore 452010, Madhya Pradesh, India

ARTICLE INFO

Keywords:
Machine learning
Ensemble
Simple Logistics
Naive Bayes
Vote.

ABSTRACT

Ensemble methods give better performance compared to a single machine learning algorithm. *Vote* is one of the best ensembles. Vote merges predictions from Simple Logistics and the Naive Bayes algorithms in the present work. The paper presents a new ensemble approach – Ensemble Particle Swarm Optimization (EnsPSO). The EnsPSO approach is a combination of (i) Vote, (ii) Correlation based Feature(s) Selection (CFS) method, (iii) PSO algorithm and (iv) random sampling method. The EnsPSO shows better performance results than Vote. The EnsPSO shows higher classification accuracy (96%) as compared to Vote (84%). The performance enhancement of EnsPSO is also proved using ten-fold cross validation on 3 standard datasets.

1. Introduction

Machine learning algorithms enable effective decision making (Wolpert, 1992; Alpaydın, 2004; Chaudhary et al., 2016a; Chaudhary et al., 2016b; Uddin et al., 2019) when used for the cases of high dimensional agriculture data (Chaudhary et al., 2013a; Chaudhary et al., 2013b; Liakos et al., 2018; Rangarajan et al., 2018; Lawrence et al., 2020). The algorithms efficiently mine the complex relationships in the data (Rocha et al., 2010). The feature selection methods help in choosing the most relevant features from the big datasets (Timmermans and Hulzebosch, 1996; Kundu et al., 2011; Hill et al., 2014; El-Bendary et al., 2015). Researchers showed that Logistic Regression and Naïve Bayes correctly identify the plant diseases (Baker and Kirk, 2007; Gutierrez et al., 2008; Sankaran et al., 2010; Phadikar et al., 2013).

Brinjal, Beet, Cabbage, Celery, Chilli, French bean, Okra, Onion, Turnip, Potato, Tomato and Pepper are the vital vegetable crops. An important reason for their unstable and less production is the incidence of pest infections and diseases. Different bacteria, fungi, viruses, nematodes and physiological disorders are responsible for diseases. Exact recognition of disease is a multiclass classification problem.

Present work is conducted for classification of diseases using data samples for Anthracnose, Bacterial wilt, Black leg, Black-rot, Chilli-mosaic, Club-root, Downy-mildew, Early blight, Fusarium wilt, Gray mold, Late blight, Leaf-spot, Onion-smut, Powdery-mildew, Rust, Septoria leaf spot, Verticillium wilt, Yellow vein mosaic.

The ensembles classify better than the individual machine learning algorithms (Hansen and Salamon, 1990; Schapire, 1990; Breiman, 1996; Ho, 1998; Bay, 1999; Opitz, 1999; Ting and Witten, 1999; Zhang and Webb, 1999; Dietterich, 2000; Stamatatos and Widmer, 2005; Kotsiantis, 2007; Sun et al., 2007; Bolón-Canedo et al., 2012; Farid et al., 2014).

The present work suggests a new EnsPSO approach with intent to enhance the performance outcomes of Vote. The EnsPSO is a combination of (i) Vote, (ii) CFS method, (iii) PSO algorithm and (iv) random sampling method. The work also presents performance comparison of newly proposed EnsPSO approach with Vote ensemble. The EnsPSO approach is applied for recognition of vegetable crop diseases. Section 2 describes the details of the materials and methods used. Section 3 describes proposed EnsPSO approach. Section 4 presents the results and discusses them. Section 5 summarizes the conclusions drawn.

2. Materials and methods

Present work is conducted using WEKA (Witten and Frank, 2005; Hall et al., 2009). WEKA consists of various supervised and unsupervised machine learning algorithms. It provides an extensive set of data pre-processing and modeling methods.

* Corresponding author.
E-mail address: archana_sesty@rediffmail.com (A. Chaudhary).

https://doi.org/10.1016/j.compag.2020.105247
Received 15 January 2020; Received in revised form 24 June 2020; Accepted 22 August 2020

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

Crop Disease Recognition using Machine Learning Algorithms

Archana Chaudhary Thakur

Abstract: Classification is a method of observing the features of a new object and assigning it to a known class. Machine learning classification problem consists of known classes and a vivid training set of pre-categorized examples. The work diagnoses groundnut diseases using outstanding machine learning algorithms namely simple logistic, decision tree, random forest and multilayer perceptron for accurate identification of groundnut diseases. Experiments are conducted with the help of 10-fold cross validation strategy. The results advocate that above mentioned classification algorithms diagnose the groundnut diseases with excellent accuracy level. Simple logistic and multilayer perceptron show outstanding performance than other algorithms and result in 96.37% and 95.80% disease classification accuracy. Random forest and decision tree algorithms provide fair accuracies in less time. These machine learning algorithms can be used in diagnosing other crop diseases also.

Keywords: Decision tree, Machine learning, Multilayer perceptron, Oilseed diseases, Simple logistic.

I. INTRODUCTION

Machine learning is a sub-field of Artificial Intelligence that deals with development and study of system that can learn from data. It deals with developing programs that learn from experience and is also a recent field of research. Machine learning methods are widely used as compared to various statistical methods as machine learning methods do not consider basic data assumption. The accurate prediction systems for identifying crop diseases widely use machine learning algorithms [5, 6, 7, 8, 9]. Appropriate recognition of groundnut diseases helps in reducing yield losses and also notifies the agriculturists to initiate competent disease prevention methods. Now day's simple logistic, naive bayes, genetic algorithms, multilayer perceptron are widely used for predictive modeling. Researchers have used machine learning algorithms for diagnosing various crop diseases [3, 4, 19]. The authors have employed logistic model, naïve bayes and random forest machine learning algorithms for recognition of oilseed diseases [3, 4] and the authors have also confirmed the results using UCI standard datasets [12]. Machine learning algorithms are also found useful for forecasting powdery mildew in mangos [15]. A classification problem in machine learning is a supervised learning problem. It consists of observing the input data (training set) and to develop an exact portrayal or model of every class with the help of attributes present in data [13]. The developed model is then used to catalog the test data with unknown class labels. The decision tree, multilayer perceptron and simple logistic machine learning classification algorithms are exercised in different applications [19]. Groundnut is the basis of valuable protein and is used widely in eatables and industries. Groundnuts are rich source of vegetable oil. Additionally groundnuts also contain major quantity of minerals, salts and vitamins. Timely recognition of groundnut diseases plays a vital role in enhancing the productivity and production of good quality crops. Clustering and Rough Set Theory (RST) based methods for leading to description of crop diseases was suggested in [10, 11]. The major focus was to show the appropriateness of the presented algorithms for developing various disease clusters and then to analyze features of a particular disease. The paper is arranged as follows: Section II consists of materials and methods used in the present work. Section III portrays results and discussions. Section IV finally presents the conclusions drawn from the present work.

II. MATERIALS AND METHODS

The real life groundnut disease dataset is used in the present work [4]. The disease classes in the dataset are Alternaria leaf spot, Charcoal rot, Collar rot, Cylindrocladium black rot, Early leaf spot, Fusarium rot, Late leaf spot, Myrothecium leaf blight, Powdery Mildew, Rust, Stem rot, Yellow mold, Zonate leaf spot. The dataset in [4] is enhanced by appending symptoms of 4 more groundnut diseases namely Anthracnose, Rhizoctonia foliar blight, Verticillium wilt and Scab. The new dataset has 17 disease classes and 2022 instances with no missing feature values. The dataset contains all the nominal disease influencing features. There are 26 disease influencing features and one feature as the disease target class representing groundnut diseases.

A. Machine Learning Algorithms

Machine learning algorithms namely decision tree, multilayer perceptron, simple logistic and random forest are used to recognize various groundnut diseases. Experiments are conducted using machine learning suite WEKA [18].

1) Decision Tree

It is a famous machine learning algorithm [17]. It is one of the best prediction ensemble algorithms used in various fields like statistics, data mining and machine learning. Decision trees use the principle of information gain for creating a decision tree from the training set. For each node the algorithm selects a feature which most promisingly partitions the data. The chosen feature has the greatest information gain. The same method is carried out until the entire decision tree is formed, having a representation of all the class examples based upon which test samples are classified.

Revised Manuscript Received on September 05, 2020.
* Corresponding Author
Archana Chaudhary*, School of Computer Science & IT, Devi Ahilya University, Khandwa Road, Indore 452001, Madhya Pradesh, India. E-mail: archana_scs@yahoo.in

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, 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.

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.

Revised Manuscript Received on September 05, 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

International Journal on Emerging Technologies 12(2): 262-265(2021)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Machine Learning Algorithms for Software Assessment

Archana Chaudhary Thakur*
Assistant Professor, School of Computer Science & IT,
Devi Ahilya Vishwavidyalaya, Indore, (Madhya Pradesh), India.

(Corresponding author: Archana Chaudhary Thakur)
(Received 24 June 2021, Revised 22 July 2021, Accepted 24 August 2021)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Software maintainability is the extent to which a software can be understood, enhanced, or modified to adapt the present environment. It is an important quality measure according to ISO standards. It has been a major concern from decades and even in present times. Object-Oriented methodologies are used to develop several open-source software. Various researchers have employed statistical techniques on metric data excerpted from software to assess software quality. Various researchers have employed machine learning algorithms for fault prediction, reliability prediction, threshold estimation, but assessing software maintainability is still a great challenge now days. Machine learning algorithms can be explored to assess software maintainability. The contribution of the present work is to use machine learning algorithms for estimating software maintainability. In the present work machine learning algorithms are used for assessing the maintainability of an open-source software. The present work also emphasizes upon the relationships amongst different maintainability metrics.

Keywords: Object-Oriented methodologies, ISO, Machine learning, Open- Source Software, Maintainability, Design metrics.

I. INTRODUCTION

The ISO standards suggest software quality using eight features. One of the vital measures for assessing a software is software quality. According to ISO standards software maintainability is one of the most important quality features since decades. Object-oriented methodologies are used for creating various open-source software today. Therefore, in order to assess the quality of any open-source software, it is necessary to assess the maintainability of open-source software. When software applications are designed, the design metrics are attained. Design metrics can be employed as essential indicators to assess software maintainability [12]. Machine learning is a field of artificial intelligence [22]. Software applications improve themselves through experience using machine learning algorithms. Machine learning emphasizes on developing predictive models [18, 19]. These predictive models are used for training various software applications. Machine learning suit symbolize a set of application software that learn from a specified set of data and attain predictions on the novel dataset based upon its learning experience [20, 21]. Machine learning algorithms are trained with the help of an example dataset in order to make predictions on novel datasets [23]. Various categories of machine learning algorithms have achieved greater influence for example supervised learning, reinforcement learning, unsupervised learning, semi-supervised learning algorithms etc.

Machine learning algorithms are used on various diverse software applications for quality assessment [13,14]. Earlier various metrics were observed for different software applications for building predictive models. Different software metrics have also attained greater influence in maintainability prediction [9, 15, 16]. Various research works have also been conducted employing some design metrics from various suites. Software maintainability was also observed for entire software using entire metric suite [16]. Some of the well-known Object-Oriented metric suites are the MOOD suite, CK suite, and the Martin suite. In the present work, complete metric suite that is Martin suite of metrics is excerpted from a famous open-source charting application j free chart employing the J Depend tool [9]. Machine learning algorithms were then employed on the excerpted dataset. The present work establishes the relationships amongst design metrics recommended by Martin [8] and maintainability.

The present work is organized as follows. The related work is presented in Section II. The methodology used is described in Section III. Section IV presents results and discussions. Section V presents the conclusions drawn.

II. RELATED WORK

It was proposed that different metrics can be used as important measures for predicting maintainability [2]. The Maintainability Index (MI) is a software metric which evaluates the maintainability of a software code [17]. MI is a code metric exercised to assess the maintainability of a software application [2]. Various research works in the past used MI metrics and had attained significant attention [3]. The usage of MI metric was also validated in various research works [10]. The MI metric can be employed to verify the maintainability of existing code and was subsequently adapted to Object-Oriented paradigm also but with some restrictions [2]. The MI metric was validated and it was used practically in determining the maintainability of a software application [1]. It was also suggested that MI can be used to assess the quality of any source code. Machine learning predictive models were developed using design metrics and has attained great importance because of the benefits it offers. The designers can suggest changes in software design or code using machine learning predictive models and improve the

ENABLING MOBILE TECHNOLOGY SECURITY ISSUES

Archana Chaudhary Thakur*

*School of Computer Science and IT, Devi Ahilya Vishwavidyalaya, Indore – 452001, Madhya Pradesh, India.

Abstract: *There are various recent advancements in technology nowadays. Mobile technology is indispensible part of our day to day life. The extensive usage of mobile phones lead to modern ways we could never think of before like easy access to the internet, official or personal information. But with it several security risks arrive like compromised privacy or some unauthorized access to mobile phones, permit easy scope of malware developers and hackers. The Android environment is a good example that has very less limitation which leads to a less secure system for the end users. So the first focus should be to reduce these security flaws and permit users with a good experience without bothering about these problems. The paper presents vital mobile security requirements and issues. The present work also presents mobile security enhancement measures.*

Keywords: **Authenticity, Integrity, Mobile Security, Privacy.**

1. INTRODUCTION

Mobile security issues are of prime concern today as organizations depend widely on mobile technology. There exists more than 86 million and only approximately 10% are nearly secured [3]. The security concern of mobile devices has become a hot topic for discussion, but on the contrary the solutions for the same are really few. The Android environment is the most widely used operating system today at various places like TV, smart-phones, watches, vehicles in car technology etc. The main reason for the popularity of android system is due to its open source code licenses. There are over 2.1 Million applications in the Play Store but the security concern for those applications is that these are loosely secured [1, 2]. Security issues are important as people have good access to the internet nowadays from mobile phones at the same time. Security measures may include the use of biometric systems like fingerprint recognition, iris scan etc, besides passwords, pattern lock, and pin [3].

2. RELATED WORK

Mobile devices are important means of communication transfer and the most widely used technology for correspondence in the whole world. With the quick development of mobile technology, ruptures in the framework security and occurrences of exchange misrepresentation are increasing. Hence, there is a need to design a profound security framework. Modern mobiles give heaps of the capabilities of customary (PCs) and provide a significant determination of alternatives available, for example, IEEE 802.11, Bluetooth, GSM, GPRS, UMTS, and HSPA [4]. The aforementioned engaging highlights presented have resulted in a far-reaching dissemination of mobile phone that, consequently, is currently a perfect focus for assailants. Earlier mobile phones came up bundled with institutionalized working framework (OS). Less heterogeneity in operating system allowed assailants to misuse only a specific helplessness to hit a large number of different types of gadgets by causing notable security flare-ups [5]. Some instances of the risks issues with mobile phones are:

- Data leak
- Unintentional revelation of data attacks on deactivated devices
- Attacks through different Spywares
- Network spoofing related attacks
- Attacks related to surveillance systems

CERTIFICATE

(IDEA - 2K21)
08-09 October, 2021

3rd Springer International Conference
on
Data, Engineering & Applications
(IDEA - 2K21)

This is to certify that Prof/Dr/Mr/Ms Archana Thakur From School of Computer Science & IT, DAVV, Indore has participated in the 3rd Springer International Conference on Data Engineering & Applications jointly organized by School of Information Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya (The State Technical University of Madhya Pradesh) and Oriental Institute of Science & Technology, Bhopal on 08th and 09th October 2021 and presented a paper titled Study of Machine Learning Classifier for Intrusion Detection System

Dr Rajesh K Shukla
Program Chair

Dr Jitendra Agrawal
Program Chair

Prof Sanjeev Sharma
General Chair

Prof Sunil Kumar
Vice Chancellor, RGPV

Springer

IDEA