IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Machine Learning Algorithms for Intelligent Mobile Systems

Archana Thakur[1*], Ramesh Thakur[2]

[1] School of Computer Science and IT, Devi Ahilya University, Indore, India
[2] International Institute of Professional Studies, Devi Ahilya University, Indore, India

[*]*Corresponding Author: archana227@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 19/Jun/2018, Published: 30/Jun/2018

***Abstract*—** Machine Learning field has evolved from the field of Artificial Intelligence, in which machines aim to follow intellectual capabilities of human beings. Machine learning is a stream of study of computer algorithms that outperform through experience. These algorithms help in discovery of rules and patterns in sets of data. The paper presents architecture of a mobile intelligent system. It also presents machine learning algorithms useful for mobile intelligent system. Besides it also discusses performance indicators of machine learning algorithms for mobile intelligent system.

***Keywords*—** Machine Learning, SVM, CBR, CART

I. INTRODUCTION

Machine learning is an evolving field of computational algorithms that are used to imitate human intelligence by learning from the surroundings. It is considered as an important field that deals with big data analysis.

Modern developments in the theory and learning algorithms form the foundation of machine learning field. Many successful applications of machine learning algorithms have been designed and applied in different fields, ranging from pattern recognition programs, computer vision to information-filtering applications, such as the program trained to sense fraudulent credit card transactions, data mining programs, users reading preferences, self-governing vehicles which can drive on public highways, spacecraft engineering, computational Biology, agriculture, malware detection [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 26].

Mobile technology has fostered development with the help of advancing mobile terminals and mobile networks. Machine learning is useful for several mobile applications like activity recognition through sensors, categorizing different text, identifying various malwares, natural language understanding etc. Machine learning algorithms outperform in domains of Natural Language Processing, diagnostics, bioinformatics, design of internet search engines, mobile based learning, classification of DNA sequences, named entity recognition, computer vision, sequence learning, game playing applications etc.

The work presents machine learning algorithms useful for intelligent system designed for mobile devices. It also discusses architecture of intelligent system designed for mobile devices along with the performance indicators.

II. ARCHITECTURE OF INTELLIGENT SYSTEM

The success of mobile applications led to the advancement of mobile technology and services. The new mobile technological advancements serve as baseline to development of new mobile applications and related areas. Machine learning applications for mobile devices include intellectual game playing, medical diagnosis applications, application for language translation, intelligent agents etc. The followings are features of mobile applications:

- It has too less start up time as users use mobile devices for small span of time.
- It has good sensitivity and acknowledges instantly when an action is executed.
- It is designed with focused intention. It has apparently defined task set that has to be done with smallest clicks or tabs.

Additionally a mobile application has a set of style instructions for the device executing the application.

An intelligent system is a system that executes those functions which usually require human intellect. An intelligent system developed for a mobile application is called Mobile Intelligent System (MobIS). The system is helpful in effective decision making, useful in forecasting or prediction. Figure 1 shows simple architecture of MobIS. Architecture of MobIS

using sensor technology is designed in [25]. The authors in [25] incorporated key functions of data gathering, processing and transmission of information. Besides multi criteria methods were utilized to discover the best path between the source and target of MobIS.

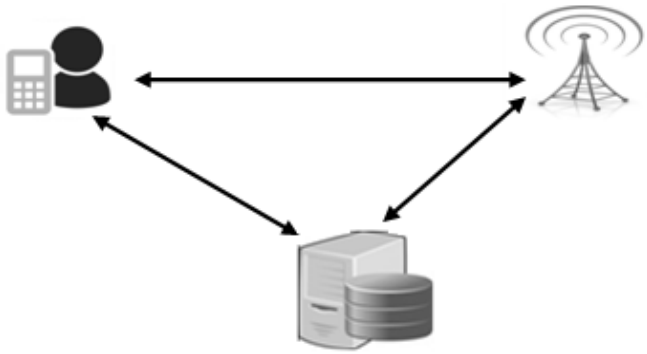

Figure 1. Architecture of MobIS

The success of every MobIS is highly dependent on the performance of learning algorithms applied on mobile applications. The indicators for estimating the performance of machine learning algorithms for MobIS are as follows:

a) Classification accuracy

b) Sensitivity

c) Specificity

Accurate evaluation of results for a known set of data is known as classification accuracy whereas correct classification of positive instances is called as sensitivity measure of a machine learning algorithm. Correct classification of negative instances determines the specificity measure of a machine learning algorithm. A machine learning algorithm achieves good quality if classification accuracy, specificity values are good and sensitivity measure is fair enough. Hence, the selection of the most suitable algorithm plays an important role in successful design of MobIS.

III. MACHINE LEARNING ALGORITHMS FOR MOBIS

Machine learning applications are in different fields of Computer Science, AI, Probability and Statistics, Information theory and Psychology. Machine learning classification algorithms that are useful for MobIS are described in this section.

A. Support Vector Machine (SVM)

It is a classification algorithm suggested by Vladimir Vapnik and colleagues at AT&T Bell laboratories in 1992. It consists of learning algorithm that analyze data and recognizes patterns. It is useful for both classification and regression tasks. The core idea of SVM is use of a nonlinear mapping function that converts data in input space to data in feature space in a way to make a problem linearly separable [27]. The input to SVM is a set of training samples. Each marked sample belongs to one of the two sets or vectors – positive samples or negative samples. The training set of an SVM algorithm makes a representation that allocates novel samples into one vector (positive samples) or another (negative samples). It generates a separating optimal hyper plane in the space considering input as two groups of vectors in an n-dimensional vector space. It attempts to maximize the margin between the two given sets or groups [11]. The working of SVM is shown in Figure 2.

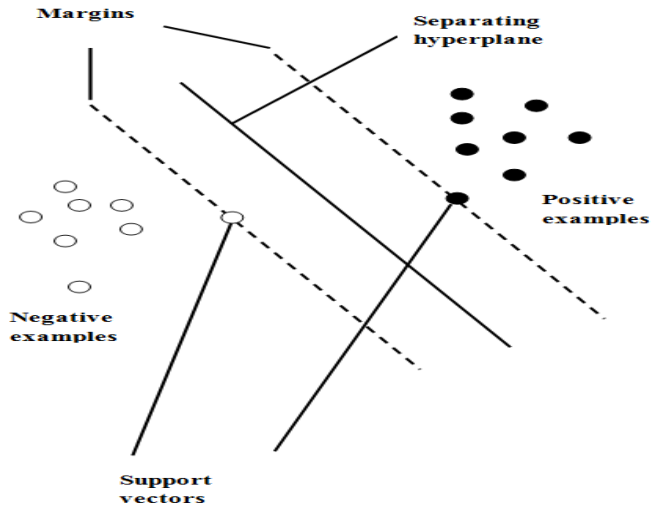

Figure 2. Working of SVM

Novel samples are consequently arranged into that similar space as expected in a group based on the area of the margin they are associated. SVM has a excellent theoretical foundation and is found very useful in real world applications. SVM is useful in predicting DNA sequences in bioinformatics, designing diagnostic applications, protein structure formation etc.

B. Decision Tree

It is a tree or hierarchical structure based classification algorithm [12]. A tree has one root, branches, internal nodes and leaves. It is a graphical structure having internal nodes that signify tests on one or more characteristics and leaf nodes signify the decision results. Figure 3 shows a simple decision tree.

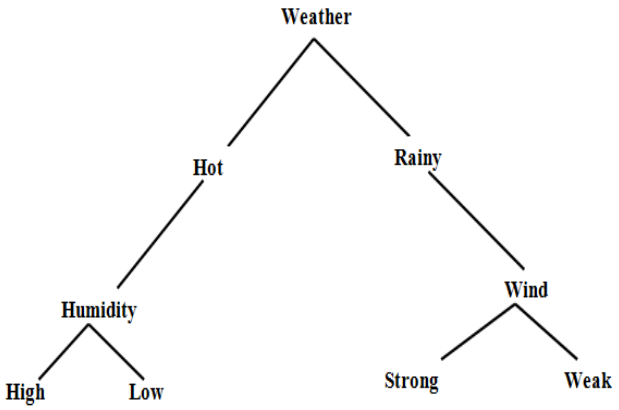

Figure 3. A simple Decision Tree

In classification and prediction Decision Tree is used as a powerful algorithm. It applies various decision rules on input for forecasting or for future course of actions. It is a classification algorithm which uses principle of partitioning by making recursive partitions of instance space. It represents rules, in human understandable form in knowledge systems. Figure 4 illustrates the working model of a Decision Tree

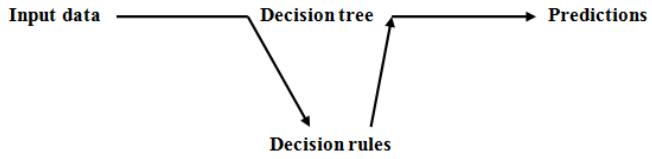

Figure 4. The working model of a Decision Tree

Decision Tree design is either from left to right or beginning from the top to bottom nodes. The internal node is divided into two or more branches. A node in a Decision Tree symbolizes a certain feature/characteristic and the branches represent decision outcomes. The Decision Tree learners discussed in the work are:

1) REPTree

It is a learner that comes under the category of a Decision Tree. It creates a Decision Tree based on information gain as the basis of division. It minimizes the tree using pruning method of reduced error [13]. The pruning method is applied till any more pruning decreases the classification accuracy. It is an efficient learning algorithm.

2) Random Tree

The Random Tree is designed at random from set of possible trees where each tree in the set has the same probability of being sampled. Relevant features are selected on the basis of random features. Random Trees are efficient and accurate models of learning.

3) J48 Tree

J48 is based on the feature values of the existing training set of data [14]. It functions on recognition of an attribute or feature that distinguishes various instances more accurately. J48 is slightly modified decision tree. J48 results in Decision Tree for specified dataset partitioning data samples recursively. It comes under the category of depth-first strategy. It selects a test from all possible tests of dividing data samples with best information gain value.

4) Decision Stump

It is also termed as single-level Decision Tree where division at the root is based on a particular attribute/value pair. It consists of single internal node (the root) that is attached to the leaves. It makes a forecast on basis of just a single input characteristic. Decision Stumps are also known as 1-rules [15]. These are also used as vital learner modules with Boosting and Bagging algorithms.

5) Random Forest

It is an ensemble learning algorithm consisting of tree-structured classifiers [16]. Each tree in forest donates a unit vote for predicting the most probable class label for each input feature. It is also useful for identifying non linear data patterns. One of the important benefits of Random Forest is that it does not show over fitting, even if greater number of trees is added to forest.

It is one of the best machine learning algorithms which is useful for different types of classifications [17]. It is an efficient classification algorithm and is robust to noise. It can handle numerical and categorical data [18]. It successfully solves many disease diagnosis problems in the literature [19, 20, 21, 22].

C.Case Based Reasoning (CBR)

It is a famous algorithm that uses solutions of the problems solved previously, for solving new problems. By using previous solution examples it then attempts to resolve step by step fresh problems. Every solution is called as a case [23].

A novel case is stated by preliminary portrayal of the problem. Each case is chosen from an arrangement of past cases and the past case is then joined with the new case by a reuse methodology. The solved case becomes an example solution to the problem characterised. The known solution is implemented sensibly to solve the real world problems after it is tested. The test phase is termed as amendment to the problem.

The retaining phase begins after testing phase concludes. The retaining phase retains vital experience for reusing it in future and case base is improved by a recently discovered case or by updating of some subsisting cases.

D.Correlation And Regression Tree (CART)

CART is used for evaluating predicting class variables values (Classification) and continuous dependent variables (Regression). Both regression and classification problems can be solved using CART [24]. CART works by designing trees and selecting an optimal tree which suits the requirements.

It needs less input for further investigation. It is the greatest advantage of this algorithm as compared to other algorithms which require more inputs and a detailed examination of intermediary outcomes and explanation of outcomes.

E.Rule Induction

It is a widely used algorithm. It is a simple algorithm as rules are simple to frame and infer than a regression model or an ANN. It works on if-then rules, decision trees or related kind of knowledge arrangements. This algorithm progresses by rule condition matching process and finds the first rule whose condition matches the example, employing an all-or-none match process [24]. It makes use of Greedy search method. Propositional learning and Relational learning are two ways of rule induction.

F.Genetic Algorithms and Genetic Programming

Genetic Algorithms and Genetic Programming are a type of computing paradigms that are useful in problem solving. In Genetic Algorithms biological evolution algorithms are applied on the principle of like natural selection [24]. The foundation of Genetic Algorithms is genes, chromosomes and population. The basis of Genetic Algorithms is Darwin theory "the survival of the fittest". The algorithm uses a fitness function which results in the best solutions selected from a set of individuals. The individuals that are found appropriate have better chance for their selection. On completion of selection process new individuals are created. This is known as mutation or crossover.

A combination of the genetic make-up of the two solution candidates generates new individuals, in crossover process. Some randomly chosen segments of genetic information are changed in order to obtain a new individual, in mutation process. The process of generation of individuals does not conclude until one of the conditions for example minimum criterion is satisfied or the desired level of fitness is attained or a specific number of generations are obtained or any of the combinations of the above is reached.

It is difficult to find out that one learning algorithm outperforms the other, though the performance of any machine learning algorithm highly depends upon the application domain and corresponding requirements.

IV. CONCLUSION

The paper discussed architecture of MobIS and performance indicators of machine learning algorithms for MobIS namely classification accuracy, specificity and sensitivity. Classification accuracy and specificity values should be relatively good whereas sensitivity should be fair enough for achieving a better performance. The paper also presented machine learning algorithms useful for MobIS. It is difficult to state that one learning algorithm performs better than the other, though the performance of a machine learning algorithm highly depends upon the application domain and corresponding requirements.

REFERENCES

[1] Damopoulos, D., Menesidou, S.A., Kambourakis, G., Papadaki, M., Clarke, N., Gritzali, N., *"Evaluation of Anomaly-Based IDS for Mobile Devices Using Machine Learning Classifiers"*, Security Comm. Networks, Vol. 5, Issue. 1, pp. 3-14, 2011.

[2] Firdausi, I., Lim, C., Erwin, A., Nugroho, A.S. *"Analysis of machine learning techniques used in behavior- based malware detection"*, In: IEEE International Conference on Advances in Computing, Control, and Telecommunication Technologies, PP. 201-203, 2010.

[3] EI-Bendary, N., EI-Hariri, E., Hassanien, A.E., Badr, A.. *"Using machine learning techniques for evaluating tomato ripeness. Expert Systems with Applications"*, Vol. 42 Issue. 4, pp. 1892–1905, 2015.

[4] Phadikar, S., Sil, J., Das, A.K. *"Rice diseases classification using feature selection and rule generation techniques"*, Computers and Electronics in Agriculture, Vol. 90, Issue. C, pp. 76–85, 2013.

[5] Kundu, P.K., Panchariya, P.C., Kundu, M. *"Classification and authentication of unknown water samples using machine learning algorithms"*, ISA Transactions, Vol. 50, Issue.3, pp. 343-520, 2011.

[6] Sankaran, S., Mishra, A., Ehsani, R., Davis, C., *"A review of advanced techniques for detecting plant diseases"*, Computers and Electronics in Agriculture, Vol. 72, Issue. 1, pp. 1-13, 2010.

[7] Timmermans, A.J.M., and Hulzebosch, A.A., *"Computer vision system for on-line sorting of pot plants using an artificial neural network classifier"*, Computers and Electronics in Agriculture, Vol. 15, Issue.1, pp. 41–55, 1996.

[8] Butler, D.R., Wadia, K.D.R., Jadav, D.R., *"Effects of leaf wetness and temperature on late leaf spot infection of groundnut"*, Plant Pathology, Vol.43, Issue. 1 pp. 112–120, 1994.

[9] Chaudhary, A., Kolhe, S., Raj Kamal, *"A Hybrid ensemble for classification in multiclass datasets: An application to Oilseed disease dataset"*, Computers and Electronics in Agriculture, Vol. 124,pp. 65-72, 2016.

[10] Chaudhary, A., Kolhe, S., Raj Kamal, *"An improved Random Forest Classifier for multi-class classification"*, Information Processing in Agriculture, Elsevier Journal Information Processing in Agriculture, Vol. 3, Issue.4, pp. 215-222, 2016.

[11] Pujari, J.D., Yakkundimath, R., Byadgi, A.S., *"Classification of fungal disease symptoms affected on cereals using color texture features"*, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 6, Issue. 6, pp. 321-330, 2013.

[12] Quinlan, J.R., *"Generating Production Rules from Decision Trees"*, In: Proceedings of the Tenth International Joint Conference on Artificial Intelligence, pp. 304–307, 1987.

[13] Zhao, Y., and Zhang, Y., *"Comparison of decision tree methods for finding active objects"*, Advances in Space Research, Vol. 41, Issue. 12, pp. 1955-1959, 2008.
[14] Quinlan, J.R., *"Generating Production Rules from Decision Trees"*, In: Proceedings of the Tenth International Joint Conference on Artificial Intelligence, pp. 304–307, 1987.
[15] Iba, W., and Pat, L., *"Induction of One-Level Decision Trees"*, In: Proceedings of the Ninth International Conference on Machine Learning, Aberdeen, Scotland, San Francisco, CA: Morgan Kaufmann, pp. 233–240, 1992.
[16] Breiman, L., *"Random Forests"*, Machine Learning, Vol. 45, Issue. 1,pp. 5–32, 2001.
[17] Seera, M., and Lim, C.P., *"A hybrid intelligent system for medical data classification"*, Expert Systems with Applications, Vol. 41, Issue. 5,pp. 2239–2249, 2014.
[18] Titapiccolo, J.I., Ferrario, M., Cerutti, S., Barbieri, C., Mari, F., Gatti, E., *"Signorini, M.G., Artificial intelligence models to stratify cardiovascular risk in incident hemodialysis patients"*, Expert Systems with Applications, Vol. 40, Issue. 11, pp. 4679–4686, 2013.
[19] Azar, A.T., Elshazly, H.I., Hassanien, A.E., Elkorany, A.M., *"A random forest classifier for lymph diseases"*, Computer Methods and Programs in Biomedicine, Vol. 113, Issue. 2,pp. 465–473, 2014.
[20] Austin, P.C., Tu, J.V., Ho, J.E., Levy, D., Lee, D.S., *"Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes"*, Journal of Clinical Epidemiology, Vol. 66, Issue. 4, pp 398-407, 2013.
[21] Tripoliti, E.E., Fotiadis, D.I., Argyropoulou, M., Manis, G., *"A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data"*,. Journal of Biomedical Informatics, Vol. 43, Issue. 2, pp. 307–320, 2010.
[22] Ramírez, J., Górriz, J.M., Segovia, F., Chaves, R., Salas-Gonzalez, D., López, M., Álvarez, I., Padilla, P., *"Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification"*, Neuroscience Letters, Vol. 472, Issue. 2, pp. 99-103, 2010.
[23] Singh, Y., Bhatia, P.K., Sangwan, O.M., *"A Review of studies on Machine Learning techniques"*, In: International Journal of Computer Science and Security, Vol. 1, Issue. 1, pp. 70-84, 2007.
[24] Chaudhary, A., Kolhe, S., Raj Kamal, *"Machine Learning Techniques for Mobile Devices: A Review"*, International Journal of Engineering Research and Applications, Vol. 3, Issue. 6, pp. 913-917, 2016.
[25] Oubaha, J., Habbani, A., Elkoutbi, M., *"Mobile Intelligent System (MIS) and a multi-criteria in MPLS networks"*, International Journal of Next Generation Networks (IJNGN), Vol. 2, Issue. 4, pp. 61-71, 2010.
[26] Jain, P.,Bhat, K., Kesharwani, H., Bhate, P., Khurana, K.P., *"Stock Market Analysis and Prediction using Hadoop and Machine Learning"*, International Journal of Computer Sciences and Engineering, Vol. 6, Issue. 5, pp. 578–584, 2018.
[27] Sharmila, R., Chellammal, S., "A conceptual method to enhance the prediction of heart diseases using big data techniques", International Journal of Computer Sciences and Engineering, Vol. 6, Special Issue. 4, pp. 21–25, 2018.

Authors Profile

Dr. Archana Thakur received M.Tech and Ph.D. from School of Computer Science & IT, Devi Ahilya University, Indore. She is working as Assistant Professor at School of Computer Science & IT, Devi Ahilya University, Indore. Before marriage she was Ms. Archana Chaudhary. She is involved in coordinating graduate-level and postgraduate-level training program in computer science for the university. She has published many research papers in various national and international journals & participated in many conferences. Her research areas include Artificial Intelligence, Machine learning, Data Mining and Soft Computing.

Dr. Ramesh Kumar Thakur received M.E. degree in computer engineering, and Ph.D. in computer engineering, from Devi Ahilya University, Indore. He is currently working as Associate Professor at IIPS Devi Ahilya University, Indore He is involved in coordinating graduate-level and postgraduate-level training program in computer science for the university. He has also worked as visiting professor at Indian Institute of Technology, Indore. He has published many research papers in various national and international journals & participated in many conferences. His research areas include Information Extraction, Machine Learning, Big Data Analysis.

