ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 1, Issue 8, November 2013

Performance Examination of Feature Selection methods with Machine learning classifiers on mobile devices

A. Chaudhary[1], S. Kolhe[2], Rajkamal[3]

Senior Lecturer, School of Computer Science & IT, Devi Ahilya Vishwa Vidyalaya, Indore, India[1]

Senior Scientist, Directorate of Soybean Research, Indore, India[2]

Professor, School of Computer Science & IT, Devi Ahilya Vishwa Vidyalaya, Indore, India[3]

Abstract Machine Learning is a field which deals with programming computers or mobile device that learns from experience. The field of Machine learning is a common research stream in Computer Science. Machine Learning techniques are helpful in several fields of Computer Science, Information Technology, Mobile computing, e-learning, Bioinformatics, Network Security, Agriculture and Web Document Categorization. Machine learning classification algorithms are used in Image Recognition, Pattern Recognition, Text Classification, Mobile message categorization, Mobile Image tagging applications, Mobile music selection according to user interests, Mobile learning. This work examines the Classification Accuracy of Bayesian classifier and Nearest Neighbor classifier on Mobile device with Android Environment. We present in this work the performance examination of both the classifiers with four feature selection methods of Correlation, Gain Ratio, Information Gain and Symmetrical Uncertainty.

Keywords PIMS, FIRT, k-NN.

I. INTRODUCTION

Machine Learning field is a branch of *Artificial Intelligence*, in which machines are prepared to imitate intelligent skills of humans. Machine learning is programming computers to optimize a performance criterion using example data or past experience [1]. This field employs statistical theory for designing models as its fundamental aim is to draw inference from data. The field of machine learning is widely used in medical diagnosis, pattern recognition, face recognition, Internet traffic monitoring, Internet malware detection, speech recognition, Mobile text filtering, Mobile tutor application, Mobile image and sound categorization, Mobile automatic image tagging, Mobile user location identification, surveillance, Internet Security. There are basically supervised, unsupervised, semi supervised and reinforcement type of learning. In modern machine learning three more types are added as transduction learning, inductive bias learning and developmental learning or Robot learning. In case of supervised learning a teacher or guide is available to guide the whole learning process. Machine learns from set examples provided as training dataset. All Regression and classification problems fall under this category. But in case of unsupervised learning there is no supervisor or guide available and aim here is to search for regularities in the input data. In this type of learning aim is draw implications on the basis of input patterns. Data mining problems wok with drawing inference from input patterns or sequences. Data mining problems fall under this category. Semi supervised learning makes use of both supervised and unsupervised learning for drawing inference. Recent research field of machine learning has revealed that learning without providing any explicit training set in conjunction with leaning with training set results in considerable improvement in learning accuracy. Reinforcement learning is when learning is associated with some scalar reward or punishment. In this type of learning goal is to maximize the number of scalar rewards. Problems related to game playing, Robot path planning, traffic control systems can be designed as reinforcement learning problems. Transduction learning or transductive inference is reasoning from observed specific training to particular test cases whereas in inductive bias learning employs a set of assumptions that the learner uses to predict outputs for inputs that it has not experienced. Developmental learning or Robot learning is a field that makes use of both the streams of machine learning as well as robotics. It is the use of techniques of allowing a robot to gain new skills or adapt to its surroundings through learning algorithms.

II. LITERATURE SURVEY

Learning algorithms were used for the explanation of mobile location in [2]. Individual human travel patterns could be

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 1, Issue 8, November 2013

easily tracked by mobile phone using mathematical models but the fundamental activities that commence the movement are still in phase that requires exploration. Further it was investigated in [2] up to what extent the behavioral methods could uncover the activities that were performed by mobile phone call locations that were captured when users initiated or received a voice call or a message. Support Vector Machines, Artificial Neural Networks, K Nearest Neighbor were applied for estimating Function Points of software in [3]. The Experiments performed in [3] showed that Artificial Neural Networks and Support Vector Machines are efficient models for function point prediction. Cybercrime Detection Model was used to enhance the classification accuracy of Facebook dataset in [4]. Support Vector Machines, AdaBoostM1 and Naïve Bayes classifiers were implemented on the Facebook dataset. The experimental results in [4] showed that Support Vector Machine is efficient and effective as compared to AdaBoostM1 and Naïve Bayes. Further the accuracy of Support Vector Machines was improved by applying different kernels. The ability of a mobile user for English learning was enhanced by a Personalized Intelligent Mobile Learning System (PIMS) designed in [5]. PIMS recommended English news articles to learners based on the learner's reading abilities and was evaluated by the proposed fuzzy Item Response Theory (FIRT). PIMS has been successfully implemented on the Personal Digital Assistant (PDA) to purvey personalized mobile learning for encouraging the reading ability of user. The Experimental results in [5] indicated that the proposed system supplies an effective and efficient mobile learning mechanism. Machine learning techniques were applied to dynamic video adapting; wireless ad-hoc networks and multi hop broadcast protocol in [6, 7, 8].

III. FEATURE SELECTION METHODS

In order to decrease the complexity of machine learning algorithms feature selection algorithms are used. Feature selection methods find out important features and eliminate unimportant, redundant or noise based features to reduce the dimensionality of the feature space. Feature selection improves efficiency, classification accuracy of the models developed by learning algorithms. There are two filter methods in machine learning namely feature selection and feature extraction. The aim of feature selection algorithms is to find features that result in best classification and eliminate the other irrelevant features whereas feature extraction methods search for a new set of dimensions that are combinations of original dimensions. There are many feature selection methods.

A. Correlation Based Feature Selection method

This method employs use of heuristic for evaluating the value or rank of a subset of features. This heuristic works upon the worthiness of individual features for anticipating the class labels as well as the level of inter correlation among them [9]. This method employs hypothesis ''A good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other'' as in [9]. This method ignores the unimportant features if their correlation with the class is weak. Redundant features are also undesirable once they are highly correlated with one or more of the other features. The subset evaluation function used in this method is given by

$$Merit_s = \frac{knum\ \overline{r_{cf}}}{\sqrt{knum + (knum - 1)\overline{r_{ff}}}} \quad (1)$$

where knum represents the number of features within a given subset s, r_{cf} is the feature–class correlation mean, r_{ff} is the average feature–feature inter-correlation and $Merit_s$ represents the merit of S.

B. Information Gain

Information Gain is an important metric used for ranking features. It works upon the intensity of information gained for classification as long as the feature is considered. This method of feature selection evaluates the quantity of impurity in a group by finding out entropy. This method is popular method for machine learning classification problems. Information Gain is calculated by the feature's influence on decreasing overall entropy [9, 10]. It calculates the change in information entropy given a feature as

$$Info - Gain(Ch, Feature_i) = H(Ch) - H(Ch|Feature_i) \quad (2)$$

where

$$H(Ch) = -\sum_{c \epsilon Ch} p(c) log_2\, p(c) \quad (3)$$

$$H(Ch|feature_i) = -\sum_{f \epsilon Feature} p(f) \sum_{c \epsilon C} p(c|a) \log_2 p\,(c|a) \quad (4)$$

using this simple measure, an ordered set of features may be obtained depending upon the information they provide for classification.

C. Gain Ratio

Gain Ratio method of feature selection is somewhat similar to Information Gain. It measures the gain in information for classification related to entropy of the given feature. In other words it investigates the merits of an attribute or a feature by computing the gain ratio with corresponding to the class.

$$Gain\_Ratio(Ch, Feature_i) = \frac{H(Ch) - H(Ch|Feature_i)}{H(Feature_i)} \quad (5)$$

where H(Ch) represents the entropy of class Ch, $H(Ch|Feature_i)$ represents entropy of class Ch given $Feature_i$ and $H(Feature_i)$ is the entropy for $Feature_i$.

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 1, Issue 8, November 2013

D. Symmetrical Uncertainty

Feature Redundancy refers to feature correlation. Two or more features are redundant to each other if their values are completely correlated [15]. For the features that show non-linear correlation Symmetrical Uncertainty measure is commonly used. Symmetrical Uncertainty is defined as

$$Sym - Unc = 2\left[\frac{Info - Gain\,(R|S)}{H(R) + H(S)}\right] \quad (6)$$

where

$H(R) =$

$-\sum_i P(x_i)\log_2 p\,(xi)$ *is the entropy of a variable R and* $in(R|S) =$

$HR - H(R|S)$ is the information gain from R provided by S. This method is not a better choice when one feature is correlated with a set of features.

IV PERFORMANCE MEASURES FOR MACHINE LEARNING ALGORITHMS

The performance of any Machine learning algorithm can be verified by some simple measures. A classification is correct if it can be checked by computing the number of correctly classified class examples (true positives), the number of correctly identified examples that are not members of the class (true negatives) and examples that either were incorrectly awarded to the class (false positives) or that were not identified as class examples (false negatives) [11]. These four elements form a confusion matrix for binary classification

$$Confusion\ Matrix\ for\ classification = \begin{bmatrix} t_P & f_n \\ f_p & t_n \end{bmatrix} \quad (7)$$

A. Accuracy: It is also called Classification Accuracy. It is a simple measure to evaluate a classifier. It is defined as the degree of correct classifications of a model. It can also be measured in percentage.

B. Precision: It is the number of correctly classified positive examples with respect to the number of examples that exist in the system as positive. Precision for true positive (t_p) and false positive (f_p) is calculated by

$$Precision = \frac{t_p}{t_p + f_p} \quad (8)$$

C. **Recall**: It is the number of accurately classified positive examples divided by the number of positive examples in the data. Recall also measures of sensitivity. Precision for true positive (t_p) and false positive (f_p) is calculated by

$$Recall = \frac{t_p}{t_p + f_n} \quad (9)$$

V. CLASSIFICATION ALGORITHMS

A. Bayesian Network

Machine learning classification problems deal with allocating a class or category to an instance or observation [9]. A classifier classifies on the basis of a training set that consists of many correctly classified examples. These are known instance and class pairs. Every instance is related by a number of features and the classifier maps this set of features into a particular class. A Bayesian Network is used to represent succint form of probability distributions. It is a graphical model of a joint multivariate probability distribution that expresses properties of conditional independence between variables [12]. A Bayesian model can be used to represent the probabilistic association between crop diseases and symptoms. The network can thus be used to calculate the probabilities of the existence of various crop diseases, given the symptoms. A Bayesian model works on Bayes rule that can be stated as

$$P(Hypothesis|Data) = \frac{P(Data\,|\,Hypot\ hesis).P(Hypot\ hesis)}{P(Data)} \quad (10)$$

where P(Hypothesis) is prior probability of hypothesis, P(Data) is prior probability of Data, P(Hypothesis | Data) is the posterior probability, P(Data | Hypothesis) is the likelihood.

B. Nearest Neighbor

Nearest neighbor learning is also termed as k-Nearest Neighbor (k-NN). It is a non-parametric type of classification algorithm. It is also called Instance Based learning algorithm. This method predicts object values or class membership on the basis of k closest training examples in the feature space. It is a popular method for classification and regression. It is a lazy learner where function is approximated locally and all computation is postponed until classification task is complete. The training set supplied are vectors in a multidimensional feature space where each instance is associated with a class label. The training stage of this algorithm consists of only storing the feature vectors and corresponding class labels of training examples. During classification stage, an unlabeled vector or a test point is classified by assigning the label which is most frequent among the k training examples nearest to that unlabeled vector or test point, where k is any user defined constant. A commonly used distance measure in case of continuous variables is Euclidean distance. In case of discrete variables Hamming distance is used as a distance measure.

VI. EXPERIMENTS AND RESULTS

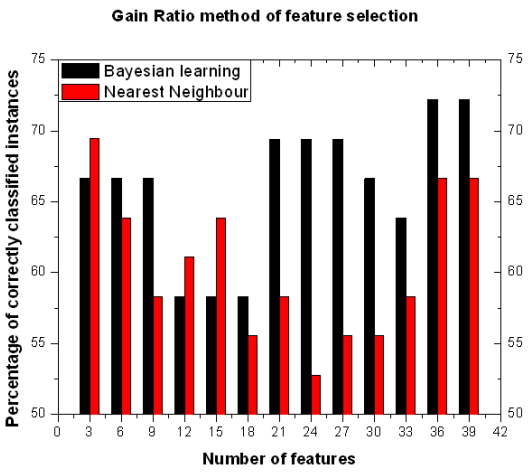
Both Bayesian as well as Nearest Neighbor classifier were used in this work with Indian crop variety dataset [13]. The four feature selection methods of Correlation, Gain Ratio, Symmetrical Uncertainty and Information Gain were used to evaluate the Classification Accuracy. This dataset contained 39 features with 200 instances. There were 75 crop varieties of India. Feature selection algorithms were applied to rank

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

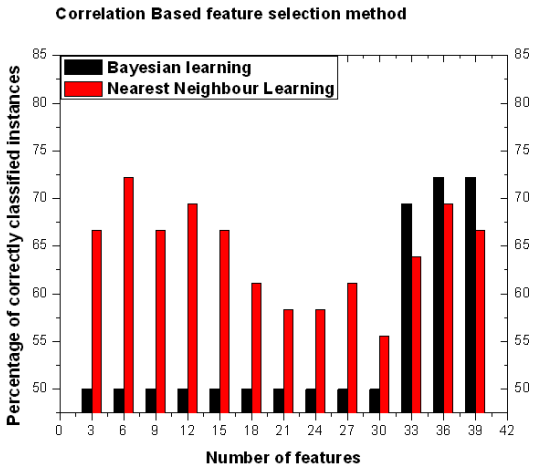
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 1, Issue 8, November 2013

crop features. Feature extraction algorithms were applied for extracting important features and discarding unimportant features. Feature space was divided into three feature subsets. Every time three new features were added to the selected features. Ten consecutive runs were performed by including the features from best ranked feature subset, average ranked feature subset and below average ranked feature subset. Experiments were performed using [14] on mobile device in which several new metrics and enhancements were added that do not exist in the current distribution. Experiments were validated using 10 cross folds validation in which the entire dataset was fragmented into 10 equal sized subsets and classifier was trained on 9 subsets and tested on remaining subset. The graphs below show the performance of classifiers

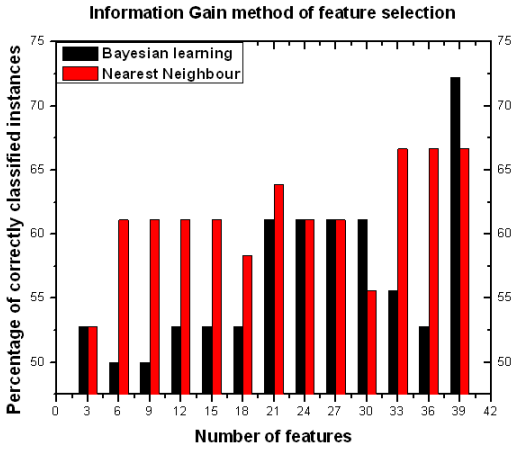
GRAPH I



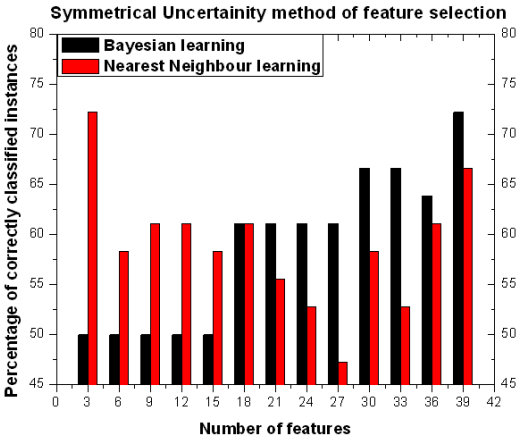
GRAPH II



GRAPH III



GRAPH IV



The table below shows the average percentage classification accuracy of Bayesian and Nearest Neighbor classifier with four feature selection methods.

TABLE – I AVERAGE CLASSIFICATION ACCURACY

Feature selection method	Bayesian classifier Average Classification Accuracy (in percent)	Nearest Neighbor classifier Average Classification Accuracy (in percent)
Correlation method	50	63.6
Gain Ratio method	64.99	59.43
Information Gain method	50.27	54.44
Symmetrical Uncertainty method	56.11	58.6

VII. CONCLUSION

The results showed that in case of Correlation based method of feature selection Nearest Neighbor classifier performed better classification than Bayesian classifier. In case Gain Ratio method of feature selection Bayesian method

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 1, Issue 8, November 2013

performed better classification than Nearest Neighbor classification method. In case of Information Gain and Symmetrical Uncertainty methods Nearest Neighbor classifier yielded better classification. Overall percentage Classification Accuracy with Nearest Neighbor was better than Bayesian method.

REFERENCES

1. Ethem 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, pp. 3299–3311, 2013.
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, 2(1), 2013.
5. C. M. Chen, S. H. Hsu, "Personalized Intelligent Mobile Learning System for Supporting Effective English Learning",
Educational Technology & Society, Volume 11, Issue (3), pp. 153-180, 2008.
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, pp. 365 – 370, Melbourne, Queensland, 2007.
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), pp. 1742 – 1747, 2011.
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, Volume 97, pp. 47–55, 2013.
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, Volume 45, pp. 427-437, 2009.
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), Volume 7244, pp. 254-265, 2012.
13. Directorate of Soybean research, India (www.nrcsoya.nic.in).
P. Liu, Y. Chen, W. Tang, Q. Yue, "Mobile WEKA as Data Mining Tool on Android**.** Advances in Intelligent and Soft Computing Volume 139, pp 75-80, 2012.
14. H. Liu, H. Motoda, "Computational methods of feature selection", Chapman and Hall/CRC, Taylor and Francis Group, NW, 2008.