Proceedings of the 12th INDIACom; INDIACom-2018; IEEE Conference ID: 42835
2018 5th International Conference on "Computing for Sustainable Global Development", 14th - 16th March, 2018
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Classification Techniques for Activity Recognition via Wearables

Sunita Kumari Chaurasia
IGDTUW, New Delhi, INDIA
Email: sunita.ait@gmail.com

S. R. N. Reddy
IGDTUW, New Delhi, INDIA
Email Id: rammalik@yahoo.com

***Abstract*–Body worn devices are augmenting the capabilities of human being, also known as wearables. What if the device is aware of the context of the user? Now day's wearables are being used in recognizing the activity performed by the wearer so that the device can work better by knowing users context. Recognizing context is not an easy task; if the numbers of class of activities are more, then accuracy is questionable. For determining the activity there are various machine learning algorithm which can be used for classification. With respect to wearable & activity recognition various classification techniques are reviewed and new method of classification fusion is proposed for improving accuracy and reducing errors.**

Keywords–Activity recognition, Classification Algorithms, Classification fusion, Wearables.

NOMENCLATURE

AR: Activity Recognition, GMM: Gaussian Mixture Model, SVM: Support Vector Machine, k NN: k- nearest neighbour, HMM: Hidden Markov Model. MLP: Multilayer Perceptron.

I. Introduction

We are living in an era where people are shorting with time and want to perform their task on the way, but they don't want to carry any device. The technological advances have also supported this scenario and give rise to always on portable computer which is un-obstructive and wearable. What if the device detects the context and work as per the context of user? Activity is also considered as one of the context. Wearables have advance sensors which can be used for activity recognition. There are various activities like running, walking, jumping, ascending and descending stairs, etc, which, if will be known to the device or any other application (which works for the assistance of the user) the assistance can be better or accurate. In this paper we have tried to find out various algorithm used for the classification of these activities with respect to wearables and proposed a new method of classification fusion.

II. Classification Algorithms for AR

Classification is defined as Data Mining or Data Analysis Techniques used to discover group membership of data instances [1], [2]. It is basically a machine learning method used in partitioning data into various classes as per some common properties. It has a two steps [1]. In first step classification model is constructed i.e learning step. It can also be viewed as mapping function, y = f (X), which determine the class label y for a given instance X.

The Second is classification step in which by using the constructed model, the class to which the given data belongs, is predicted.

The Learning can be of two types [1], [2]:

1. *Supervised Learning*: When the class label of training data are provided or the training samples are aware of their class to which they belongs.
2. *Un Supervised Learning*: When the class label of training data are unknown or the training samples are unaware of their class.

We can also put them in Parametric or Non Parametric algorithm. Parametric Algorithms make assumptions or reduce the function to some known form. No matter how much data we provide to this model, it won't change its decision [3]. Nonparametric Algorithms does not make strong assumptions about the form of the mapping function. They are good when we have vast data without much prior knowledge" [3].

As per the start of learning process we call them as eager learner or lazy learner. When a set of training tuples are given & the algorithm constructs a generalization model before the arrival of testing tuples they belongs to eager learner. These models are ready to classify unseen tuples. Lazy learner waits till the last minute, when a training tuple is provided, they just stores the data (or does only negligible processing) and waits for the test tuple to be provided. There are various machine learning algorithm used to find out the class of unseen instance. We are focusing on the algorithm used in AR:

A. Decision Tree

Decision Tree is the oldest and simplest machine learning algorithm, which uses branching method to illustrate every possible outcome of a decision [4] . It is a hierarchical flow chart like structure in which nodes, branches and leaves are present [1], [5]. Internal nodes are for testing an attribute, the outcome of the test is shown by branches and leaves symbolize the class label [6]. The decision process starts from any selected attribute node; gets continued till it reaches to the leaf node, this leaf node is the class to which that input data

Copy Right © INDIACom-2018; ISSN 0973-7529; ISBN 978-93-80544-28-1

belongs. It is a greedy algorithm where tree is constructed in a divide-and-conquer, top to down and recursive manner [7].

The aim of this process is to find the smallest possible tree. It tries to select the best attribute based on which the grouping of the objects is to be done, that would make the data Pure as possible after the split. The purity can be measured by a function called Entropy. If there are m numbers of classes, the entropy at particular node is calculated as [1]:

$$Info(D) = -\sum_{i=1}^{m} p_i \, log_2(p_i)$$

where pi, represents nonzero probability of any arbitrary tuple present in D belonging to class C_i.

$$C_i = \{1,2, \ldots . m\}$$

Information gain is used for selecting the attributes. It is calculated as:

$$Gain(A) = Info(D) - Info_A\,(D)$$

The expected reduction in the information requirement is Gain(A) due to attribute A. Highest information gain attribute, is selected for splitting the node.

B. Logistic Regression

This technique helps in finding the dependent variable's value from the independent variables. Here the variables are numeric in nature. There are variety of of regression, for exa. linear, multiple, polynomial etc. Logistic Regression is a generalized linear model, in which the dependent variable is binary or dichotomous, i.e. it only has two possible outcomes exa: true or false, pass or fail etc. The equation to predict a logit transformation of the probability of presence of interested characteristic is:

$$\text{logit (p)} = \ln(\tfrac{p}{1-p})$$

C. Support Vector Machine

It classifies linear as well as nonlinear data [1], [7], [8]. Training data is transformed into a higher dimension, and then SVM searches for optimal linear hyperplane which seperates the different classes, this hyperplane can be understand as a "decision boundary". This boundary separates the instance of one class from other class. The SVM uses the essential training tuples (support vectors) and margins to find its hyperplane. There can be infinite number of hyperplane between two classes but the algorithm try to find the hyperplane with greater margin. The training time of SVM algorithms are more, but they give more accurate results. They are less susceptible to data overfitting as compare to others. SVMs can be used for both numeric prediction and classification. For linear hyperplane formulae is [1]:

$$W.X + b = 0$$

Here, W= {w1, w2.... wn} i.e weight vector and n represents no of attributes.

Considering the Lagrangian formulation, the decision boundary or the Maximum Margin Hyperplane is shown as [1]:

$$d(X)^T = \sum_{i=1}^{l} y_i \propto_i X_i \; X^T + b_0$$

Here y_i represents class label of support vector X_i; testing input X^T; Lagrangian multipliers αi and b0 are numeric parameters which is automatically calculated by optimizer or SVM; l represents no of support vectors.

D. Naïve Bays

It is a probabilistic based technique, which learns Probability of an object with certain features belonging to particular group or class. A posterior probability of every class is calculated by estimating conditional probabilities from the training input. It is called naïve because it makes hypothesis that the occurrence of a one feature is independent of the occurrence of another feature [1], [8]. It is primarily used for text classification and best applied in filtering spam messages. It is known for its simplicity and effectiveness. Bayesian Rule is given as:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Here P (X|Y) is conditional Probability of occurrence of event X when Y is valid. P (X) and P(Y) are Probability of occurrence of event X and Y. P (Y|X) is conditional Probability of occurrence of event Y when X is valid. X is called the proposition or Class and Y is called the evidence or Predictor.

E. k-nearest neighbor

k-nearest neighbors (kNN) is a simple non parametric method which is used to classify unmarked observations by providing them a class of the most analogous marked examples [1], [5], [7] , [8]. It is a supervised lazy learner technique which learns by analogy. Here the training instances are represented in multiple dimension space. When any unknown tuple arrives, k-nearest neighbor classifier searches the k training instances, which are closest to the unknown tuple. Then it places this unknown tuple to the nearest class. These k training instances are the k "nearest neighbors" or "closest neighbors" of the unidentified tuple. This technique is sensitive to the value of k. If the value of k increases the prediction time also increases. "Closeness" can be defined as a distance metric, for exa. Euclidean distance. The Euclidean distance between two points or tuples, can be shown as [1] :

$$\text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2}$$

F. Multilayer Perceptron

It is an artificial neural network model based on feed forward or back propagation where sets of input data are mapped into set of suitable outputs [8], [9]. It is based on neural networks which imitate the properties of biological neurons available in the human brain, propagate activation signals and encode knowledge in the network links [1], [7]. A single perceprton can only solve linearly separable problem. MLP is a supervised learning technique in which multiple layers of nodes are present. It forms a structure like directed graph, where each layer is fully connected to the next layer. The formulae for calculating output at a hidden layer is:

$$y_i = f(\sum_{j=1}^{m} w_{ij}x_j + b_i$$

where w_{ij} is weight from j input node to i hidden node, x_j is input from j node, and b_i is the bias at activation node.

G. Hidden Markov Model

It is used for sequential or temporal data. In the first order Markov probability of moving to the given state depends only on the present state and in the second order it depends on the present as well as the previous state [1], [2], [10]. It represents Finite set automata. As probability of a state depends on its previous state; therefore, HMM are mainly helpful in the analysis of biological sequence data. It can be used in the prediction of next state if series of states are given. Let's assume that at every time step t the system is in a state ω(t) and it produce some (visible) symbol v(t). Let's say any state ω(t) we have a probability of emitting a particular visible state vk(t), i.e emission probability. We signify this probability P(vk(t)|ωj(t)) = bjk.

The probability that the model produces a sequence of all visible state VT is [1], [2]:

$$P(V^T) = \sum_{r=1}^{r_{max}} P(V^T|\omega_r^T)P(\omega_r^T)$$

H. Random Forest

These are an ensemble learning method used for classification [8], [1], [7]. It constructs multiple decision trees at the time of training and output the class of mean prediction or that is the mode of the classes. This algorithm works as a large collection of non-correlated tree. The input sample is divided into various different random samples and trees are formulated from each random sample. For testing a new tuple these different trees are used and predict the class of that tuple. Random Forests produces several classification trees. To find out the class of a new object of the input vector, it passes the input vector through each of the trees present in the forest. Each tree gives "votes" for some class or each tree classifies the input vector and put it into some class. The forest with the highest vote is being chosen for the classification.

I. Adaboost

Boosting is ensemble (combining diverse set of learning algorithms) algorithm for classification, here each training tuple are assigned with some weights. A series of k classifiers are being learned [1] , [5] by iteration. When classifier, Mi, is learned, the weights are updated so that the subsequent classifier,Mi+1, can "pay more attention" towards the misclassified training tuples by the previous classifier Mi. The M* is the final classifier, which integrate the votes of every individual classifier, where the weight of each classifier's vote is a function of its accuracy. Adaptive Boosting also known as AdaBoost, in which subsequent weak learners are tweaked in favor of misclassified instances classified by previous classifier. It is susceptible to outliers as well as noisy data. It is less sensitive to the overfitting problem. It is referred to as best-out-of-the-box classifier. We can unify several weak classifiers into a one strong accurate classifier.

To calculate the error rate of model Mi, weights of every tuples is summed from Di which was misclassified by Mi. That is,

$$error(M_i) = \sum_{j=1}^{d} w_j \times err(X_j)$$

Here err (Xj) is the misclassification error for instance Xj: when it is misclassified, err (Xj) is 1; else, it is 0.

III. Comparison and Analysis

A detail comparative analysis has been done, shown in Table I by considering various parameters mentioned below so that conclusion can be drawn. These parameters can be viewed as:

1. *Classification Type*: The classification can belongs to various types like Tree based regression, Instance based, Bayesian and ensembles etc.
2. *Type of Learning*: It can be Supervised or unsupervised.
3. *Type of algorithm*: It can be Parametric or Non Parametric.
4. *Mathematical Formulae Used*: Calculation done for decision making.
5. *Starting of learning method*: Eager learner or Lazy learner.
6. *Advantages & Disadvantages*

IV. Conclusion and Future Scope

The comparative analysis shows that ensemble or fusion of learning algorithm gives better accuracy and can deal with overfitting and missing data but it is complex in terms of calculation. This calculation problem can be resolved as technological advancement is increasing day by day. Adaptive learning [11] has shown better result than single classifier. Various researchers have shown that [12] [13] fusion can be done in two different ways, First, Parallel where can apply

different or same classifiers simultaneously and finding the best as per majority voting, mean or mode as present in Random Forest. Second is sequential where one algorithm is applied and according to its result the weight of input vectors are modified so that it can improve its learning as in Adaboost or different classifiers are applied one by one sequentially. New method of fusion can be done where we can combine both the method of classification fusion and can make a new classifier, where we can apply classifiers in parallel and after that sequentially, so that the combined effects can further enhance the learning.

In future we will try to implement this concept after further detail study and tools selection. Future work will also include data collection from wearable devices by using special application for android based devices, on which these methods will be applied and tested.

TABLE I. COMPARISION TABLE

Machine Learning Technique	Parameters								
	Classification Type	**Classification Name**	**Description**	**Mathematical Formulae**	**Start of Learning**	**Type of Learning**	**Type of Algorithm**	**Adv.**	**Dis. Adv**
1	Tree Based	Decision tree [8], [1], [4], [5], [6], [7], [14], [9]	Tree type structure used to make decision at every internal nodes to reach till leaf i.e class to which the data instance belongs.	$\text{Info}(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$	Eager learner	Supervised	Non Parametric	1. Easy to read and understand 2. Does not require domain knowledge 3. Can be used for more than 1 class. 4. Can handle multidimensional data 5. Good accuracy.	1. Information gain in decision trees is biased in favour of those attributes with more levels. 2. Can create over complex tree and required pruning. 3. Does not work well with smooth boundries.
2	Regression or Function based	Logistic Regression [8], [1], [9]	It is a linear model in which outcome is dependent on some independent variable. Here the outcome can only be dichotomous.	$logit\,(p) = \ln(\frac{p}{1-p})$		Supervised	Parametric	1. Linear Model thus easy to implement.	1. Can be used only for binary outcomes
3	Domain Transform	Support vector machines [8], [1], [5], [7], [9]	It uses hyperplane to separate classes of data. It uses training tuples (support vectors) and margins to find its hyperplane.	$d(X)^T = \sum_{i=1}^{l} y_i \propto_i X_i\ X^T + b_0$	Eager learner	Supervised	Non Parametric	1. Can be used for classifying both linear and nonlinear data. 2. Less prone to over fitting. 3. Accuracy is high.	1. Speed of training and testing of algo is the major issue when data increases. 2. Not very much efficient for multiclass case.
4	Bayesian	Naïve Bays [8], [1], [5], [9]	Probabilistic based learning technique used to calculates posterior probabilities for each class using estimated	$P(A\|B) = \frac{P(B\|A)P(A)}{P(B)}$	Eager learner	Supervised	Parametric	1. Can build model Fast and quick prediction. 2. Can solve both categorical and continuous	1. Assumption class conditional independence, thus less accuracy. 2. Practically dependencies exist

Machine Learning Technique	Parameters								
	Classification Type	**Classification Name**	**Description**	**Mathematical Formulae**	**Start of Learning**	**Type of Learning**	**Type of Algorithm**	**Adv.**	**Dis. Adv**
			conditional probabilities from the training set.					value attributes. 3. Require small amount of training data.	among variables.
5	Instance Based	k-nearest neighbour [8], [1], [5], [7], [9]	It is a supervised algorithm which determines the class of an unknown tuple by comparing k-nearest neighbor training tuples. The class (in majority) of tuples which are closest to the unknown sample becomes class of the sample tuple.	$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2}$	Lazy learner	Supervised	Non Parametric	1. Can model complex decision spaces with hyperpolygonal shapes. 2. It is well suited for multimodal classes	1. Computationally expensive 2. Require proper storage technique. 3. It gives slower performance when applied to large set of data.
6	Neural Network	Multilayer Perceptron [8], [1], [7], [9]	It is a supervised learning based on neural network of more than one layer, can be used to solve nonlinear problem.	$y_i = f(\sum_{j=1}^{m} w_{ij}x_j + b_i$	Eager learner	Supervised	Non Parametric	1. Used to solve non-linear problems 2. Do not make any assumption.	1. It has high computational cost. 2. It require large amount of training data.
7	Tree based Ensembles	Random Forest [8], [1], [7]	It is decision tree based learning method where different trees are formulated to vote for an input test data tuple and the class with highest vote is assigned to that test data tuple.		Eager learner	Supervised	Non Parametric	1. Highly accurate. 2. Can handle the missing value. 3. Handle overfitting. 4. Can handle large dataset with many dimensions.	1. It requires creation of large number of decision tree. 2. Computationally costly.
8	Meta Based Ensembles	Adaboost [1], [5]	AdaBoost is an ensemble learning algorithm in which subsequent weak learners are adjusted in favour of misclassified instances classified by previous	$error(M_i) = \sum_{j=1}^{d} w_j \times err(X_j)$	Eager learner	Supervised	Parametric	1. Less susceptible to the over fitting problem. 2. More accuracy.	1. Computationally expensive. 2. Require several models to get trained. 3. Sensitive to noisy data and outliers.

Machine Learning Technique	Parameters								
	Classification Type	Classification Name	Description	Mathematical Formulae	Start of Learning	Type of Learning	Type of Algorithm	Adv.	Dis. Adv
			classifier.						
9	Markov Model	HMM [1], [2], [10]	Learning technique which finds the next visible state on the basis of present state or sequence of previous states.	$P(V^T) = \sum_{r=1}^{r_{max}} P(V^T\|\omega_r^T)P(\omega_r^T$	Lazy learner	Supervised	Parametric	1. Can handle real world problem. 2. Used with time series data.	1. Three problem of HMM is Evaluation, Decoding and Learning. 2. Size of data can be an issue.

REFERENCES

[1] Jiawei Han, *Data Mining Concepts and Techniques*, Third Edition ed.: Morgan Kaufmann an imprint of Elsevier, 2012.

[2] Richard O Duda and Peter E Hart, *Pattern Classification*, Second Edition ed.: WILEY, 2016.

[3] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* , 3rd ed.: Pearson, 2009.

[4] I.Bhuvana and C.Yamini, "SURVEY ON CLASSIFICATION ALGORITHMS FOR DATAMINING: (COMPARISON AND EVALUATION)," *IJARSE*, pp. ISSN 2319-8354, August 2015.

[5] XindongWu et al, "Top 10 algorithms in data mining," in *Springer-Verlag London Limited*, 2007, pp. 1-37.

[6] Saisakul Chernbumroong, Anthony S. Atkins, and Hongnian Yu, "Activity classification using a single wrist-worn accelerometer," in *IEEE*, 2011.

[7] G.Kesavaraj and Dr.S.Sukumaran, "A Study On Classification Techniques in Data Mining," in *4th ICCCNT – 2013, IEEE – 31661*, Tiruchengode, India , 2013.

[14] Shahrukh Teli and Prashasti Kanikar, "A Survey on Decision Tree Based Approaches in Data Mining," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 4, pp. 613-616, April 2015.

[8] Oscar D. Lara and Miguel A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, vol. 15, no. THIRD QUARTER, pp. 1192-1209, 2013.

[9] Yugal kumar and G. Sahoo, "Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA," *I.J. Information Technology and Computer Science*, vol. 7, pp. 43-49, July 2012.

[10] Part of the Udacity course "Introduction to Computer Vision". (2015, Feb) youtube. [Online]. https://www.youtube.com/watch?v=5araDjcBHMQ

[11] Sneha Chandra and Maneet Kaur, "Creation of an Adaptive Classifier to Enhance the Classification Accuracy of Existing Classification Algorithms in the Field of Medical Data Mining," in *INDIACom-2015, IEEE*, 2015.

[12] Moacir P. Ponti Jr., Combining Classifiers: from the creation of ensembles to the decision fusion, 2011, Institute of Mathematical and Computer Sciences (ICMC).

[13] Michal Wozniak, "A survey of multiple classifier systems as hybrid systems," *Information Fusion, elsevier*, pp. 3-17, 2014.