Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 4, April 2017, pg.376 – 379

HAR: Human Activity Recognition to Big Data Analyze of Citizens Perception of Security

R.Sangeethapriya[1], R.Maharajaa[2], K.Selvasheela[3]
[1]Computer Science and Engineering, India
[2]Computer Science and Engineering, India
[3]Computer Science and Engineering, India
[1] sangeethapriyagokul94@gmail.com; [2] r.maharajaa@gmail.com; [3] k.selvasheela@gmail.com

Abstract: Activity recognition is one of the leading application of machine learning algorithm nowadays .It is being used in the field of biomedical engineering, game development, developing better stats for sports training etc. Data from the sensors attached to a person can be utilized to train supervised machine learning models in order to predict the activity being carried out by the person. In this paper we will be using Data available at UCI machine learning Repository. It contains data generated from accelerometer, gyroscope and other sensors of Smart phone to train supervised predictive models using machine learning techniques like SVM, Random forest and decision tree to generate a model. which can be used to predict the kind of movement being carried out by the person which is divided into six categories walking, walking upstairs, walking down-stairs, sitting, standing and laying .We will be comparing the accuracy of different models using confusion matrix and K fold cross validation. The proposed system was built using a set of big data tools in order to collect, pre-process, classify, index, and visualize data. The system is able to detect whether a tweet is related to security, which is used to present in a heat map the perception of security of a city. HAR machine learning algorithm was trained to learn to recognize security characteristics in tweets. Results show that this system is a powerful tool to visually analyze how citizens perceive security

I. INTRODUCTION

In the last decade there has been an exponential increase in the number of smart phones that are available in the market .In 2014, in India alone the smart phone users are expected to increase by more than double from existing 156 million user to 364 million.[1]. With the growing number of smart phone the amount of data that can be generated from the sensor of the smart phone is also growing . The smart phone come equipped with the various inertial sensors such as gyroscope, accelerometer. The reading of these sensors change according to the movement of the smart phone.

These smart phone have integrated seamlessly into everyday life of the people around us. People carry mobile devices around for most part of the day. This enables us to track activity of the person carrying the smart phone

using the data of the sensors of their smart phone. Activity recognition takes into account this data and uses it to produce models which can be used to predict the activity being carried out by the person.

In this paper we use the data available at the UCI machine learning repository[2] which has been modified for our research. The data will be processed through supervised machine learning algorithms to produce predictive classification models that will be used to classify physical activities of the person into six categories namely Sitting, Standing , Laying , Walking , Walking upstairs , Walking Downstairs. Various models thus created will be analyzed for accuracy using K-fold cross validation and formation of confusion matrix.

II. METHODOLOGY

Existing system: This paper presents a big data analytics system that allows government to measure citizens' perception of security from the Twitter social network. The proposed system was built using a set of big data tools in order to collect, pre-process, classify, index, and visualize data. Although there are various initiatives related to security, most of them are focused on policy mechanisms. It is not easy to find government initiatives that measure PoS. A tool focused on measuring PoS would be not much useful to governments to preventively address future citizen problems related to security.

Proposed system: The system uses Twitter information to allow users to search by means of keywords. This system has been used to solve criminal cases at some countries such as Siria and Philipinas. Data mining and data analytics techniques were used in to support decision making processes related to terrorism. Authors describe a system that allows the Police to visually analyze terrorist events. We will be using Data available at UCI machine learning Repository. It contains data generated from accelerometer, gyroscope and other sensors of Smart phone to train supervised predictive models using machine learning techniques like SVM, Random forest and decision tree to generate a model. which can be used to predict the kind of movement being carried out by the person which is divided into six categories walking, walking upstairs, walking down-stairs, sitting, standing and laying .We will be comparing the accuracy of different models using confusion matrix and K fold cross validation.

III. MODULE DESCRIPTION

DATA PRE-PROCESSING

Security Detection

Automatic Tagging

Data Analytics

The problem we are dealing with in this paper falls into the category of Classification modelling . In this process we will be creating ML predictive models, which will then be used to decide which category a particular observation belongs, predictive models are formulated using the training data. The observation are needed to Classified under six categories namely sitting , standing , laying , walking , walking upstairs , walking downstairs. We will be using supervised leaning methods to create the models. We will be using 10 fold cross validation to judge the accuracy of the models generated . R studios was used to develop models.

A. Importing Data

The data was read into R studio as csv file. The data was shuffled so that there is random distribution of the observation. Now K-fold cross validation is performed using different models.

B. K-fold cross validation technique

K-fold cross validation technique is used to test the predictive accuracy of the models. The technique involves dividing the data set in K-partitions of equal number of observations. Then the K-1 number of partition are used as training set and the other complementary partition is used as validation set or the test set. The technique is usually used to prevent overfitting of data which may produce errors when prediction are made using the model developed . In our study we used 10 fold cross validation. The data has been divided into 10 different partition then one partition is assigned as the test set . Now we develop model on the remaining 9 sets which are used as training set. The new model is then test on the test partition and its accuracy is noted. This process is repeated 10 times by taking a Different partitions as the test set and complementary as training set. Accuracy of prediction is calculated noted using Confusion matrix.

C. SVM support vector machine

SVMor support vector machine is a type of machine learning technique used to analyse data and recognise patterns. It can be used to generate Classification or Regression Models . When using SVM for classification analysis the SVM model represent the Categories of classification as points in space. If the points can be separated using linear lines then simple linear SVM is created . If the points cannot be easily differentiated using hyperplanes in a particular dimension then the points are transferred to higher Dimension using Kernel Tricks . After the points are transferred to higher Dimension hyperplanes are used to demarcate the boundaries between various categories.

D. Literal Fuzzy c-Means (LFCM/AO)

LFCM/AO, partition the data into clusters such that data-points in a cluster are more similar to each other than to data-points belonging to other clusters. It tries to optimize the objective function in Eq. (1) using alternating optimization. The number of cluster centers c for LFCM/AO is decided by the user and the initial cluster centroids for c clusters are chosen randomly. We may ease the notation of LFCM/AO in the further discussion by dropping the "/AO" notation unless clarity demands it. The LFCM algorithm is illustrated in Algorithm

E. Random Sampling with Iterative Optimization Fuzzy c-Means (RSIO-FCM)

Inspired by FCM, RSIO-FCM [16] has been designed to over-come the drawbacks of rseFCM. It is a partitioning based approach which takes into account the entire data while determining the cluster centers. In RSIO-FCM, the entire data is divided into various subsets X = {X1; X2; Xsg}

In this approach, the cluster centers for the first subset are chosen randomly from the entire data. Then, final cluster centers, and membership matrix for the first subset, de-noted as V 1 and U1 respectively, are found by applying LFCM. V1 is used as the initial cluster centers for clustering of the second subset. This procedure is repeated until all the subsets are clustered. The final cluster centers for the entire data is determined by combining the membership matrix of all the subsets and using Eq. (4). Algorithm 3, highlights the steps involved in RSIO-FCM.

IV. CONCLUSIONS

From the analysis conducted in this paper out of the 10 cross validation conducted Random Forest outperformed the SVM in 8 cases. Also the Confusion Matrix analysis concluded that the misclassification error in the Random Forest Models was far less than the SVM while differentiating between sitting and standing ,whereas it produced comparable results while classifying other activity . Whereas the Tree model produced poor result throughout. It can be clearly said that the Random Forest model can be effectively utilized when used for activity recognition using the data provided by the accelerometer and gyroscope of a Smartphone.

REFERENCES

[1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[6] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

[8] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.

[9] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.

[10] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[11] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

[12] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.

