Internet of Things Data Analytics for User Authentication and Activity Recognition

Samera Batool, Nazar A. Saqib, Muazzam.A. Khan
College of Electrical and Mechanical Engineering (E & ME)
National University of Sciences and Technology (NUST), Islamabad, Pakistan
samerabatool@yahoo.com, nazar.abbas@ce.ceme.edu.pk, muazzamak@ce.ceme.edu.pk

***Abstract*— In recent years Internet of Things has been among the hot topics of research in the field of computing and information technology. It has enabled thousands of connected devices including sensors, cell phones and daily home appliances to store data for various useful purposes. The amount of data generated by IoT devices is huge, and there is a need to analyze IoT data for prospective uses. Heterogeneity and the structure of IoT data make it a challenging task. This paper presents a frame work for IoT data analytics for creating models for user identification, user authentication and activity recognition. Focus of this study is on the data set of accelerometer sensor from internet of things perspectives to add an extra layer of security. The novelty of our approach lies in the customization of the data set, and the experiments performed for the construction of models for each individual activity and user. The dataset is collected from 19 different subjects of real world conditions performing basic activities, i.e. walking, sitting and standing. This data set is used for the authentication process without requiring any additional information. In existing studies, the results are obtained using more than one accelerometer sensor reading or a combination of gyroscope sensor and accelerometer sensor. Whereas we have used single sensing tri-axial sensor reading for activity recognition and user authentication models. These models are later verified by real time data sets which were not used in the training process. The results of the experiments show accuracy up to 93%. The results obtained by the experiments are also helpful for future research directions in the field of IoT data analytics, activity recognition and user authentication. By enhancing the accuracy and adding context aware aspects in the authentication models can lead to the significant advances in the biometric authentication process using IoT data.**

Keywords—Internet Of Things; Data Analytics; Global System for Mobile Communication; Wireless Fidelity)

I. INTRODUCTION

During the last decade or so information technology industry has progressed remarkably. The introduction of smart phones, sensors Wi-Fi and GSM technologies changed the entire scenario of communication world. Internet of Things (IoT) has enabled these smart devices to communicate and store the data. IoT connects physical and digital entities to enable the new class of services.

Sensors, smart devices and home appliances are now connected through the internet to communicate with each other and enable machine to human and human to machine interaction. These devices are producing enormous amounts of data. To handle that much voluminous data is still a vital challenge. In its nature data is semi structured and streaming data. [1].

A wide variety of devices which include sensor enabled devices, cell phones and other wearable devices can connect to the internet and share data and information. At the same time the cost of technology has declined sharply. This has made access to technology very easy and affordable for everyone. And now it's very normal for most of the people to access real time information using internet. [2].

There are several IoT visions defined by the researchers. One of them is the connection of life objects to the internet for automatically sharing data and making useful decisions. The eventual goal of IoT is to connect the real-world objects and make computers to sense, hear and see the objects of real world. It is estimated by Ericson that the number of connected devices will reach up to 50 billion in 2020. [3].

IoT connected devices will soon effect each area of our life. IoT enables machine to machine and machine to human communication for the betterment of human life. IoT devices are used not only for the sensing purpose but they also require some actions to be performed on the sensed data [4, 5]. The basic building blocks of IoT are presented in the following lines.

- User Applications: It represents interaction of people and applications for business process.
- Cloud Server: Computing systems with real time processing and storage capability.
- Connectivity: Internet access, Bluetooth ZigBee, Wi-Fi 2G,3G,4G etc.
- Gateway: Communication protocols that enable communication among sensors and the network.
- Hardware:

  Physical devices equipped with sensors and actuators [6].

One major difference of the traditional Internet and the IoT is, that traditional internet consists of people who generate data and information, whereas in IoT machines are the source and disposal of information and data. IoT devices are connected for sharing, communicating and providing useful services to humans. [7,8].

The amount of data generated by IoT devices can be very useful for analytics purpose. As the connected things increase in numbers the amount of data generated by IoT devices will also increase, as this will enhance data generation speed, data volume will become vast and data space will become global. [9].

Data produced by IoT devices has its special characteristics which makes it a challenging task for analytics. Research efforts are required in this regard to cope with the increasing demand in the era of IoT. The body movement patterns that are produced by accelerometer sensor embedded in cell phone or other IoT devices are surely of great potential to identify the users.

The main objective of this research effort is to analyze existing accelerometer sensor data analytics techniques for activity recognition and most importantly focusing on user authentication. We want to enhance the accuracy of IoT data analytics solely for biometric user authentication rather than relying on any other data sources.

In this paper, we propose a framework for IoT data analytics. It utilizes the concept of models, for each class of data we create models or in other words signature, we can verify and validate these models with real time IoT test data input. We focus on the accelerometer sensor stream data for the purpose of user authentication and activity recognition.

The rest of this paper is organized as follow. Section II presents the related work. In Section III the proposed framework is presented. Section IV presents the prototype implementation of the proposed framework and results. Finally, the paper ends up with the conclusion and future work

II. RELATED WORK

There have been many efforts for analytics of IoT sensor data. We describe few in the following lines which use accelerometer sensor data.

Accelerometer sensor embedded in cell phones is used to recognize user's activities by [10]. Five difference classifier algorithms are compared and combined to optimize the performance of activity recognition. It was estimated that by the fusion method of combining the classifiers can improve the performance up to 91 %.

Cell phone based accelerometer sensor is used for the context aware authentication of users [11]. The authors proposed to consider the position of the cell phone while accelerometer measurements i.e. in pocket or in hand. Accordingly, the algorithm is proposed for the authentication and classification is performed.30 users' data set is used for the experiments and using logistic classifier in weka they have produced user identification results for in hand data 82% and in pocket data 62.55%.

A review paper is presented by [12] for comparison of different classification techniques for activity recognition using wearable sensors. The study presents results based on the three accelerometer sensors placed on the right shank, chest and left ankle. The results declare that from the supervised classification k-NN produces high accuracy, and among unsupervised algorithms HMM gives best results.

A comprehensive survey of online activity recognition using cell phone sensors is presented by. [13]. The authors present detailed analysis of the existing studies that use cell phone based sensors data for activity recognition. Outcomes of the existing studies are presented along with their limitations.

The focus of this survey remained the sensor data collected from cell phones accelerometers for activity recognition.

In [14] smartphone accelerometer data is used for activity recognition collected by 30 subjects from the age group of 19 to 45. Six activities were considered walking, jogging, sitting, standing, stair up and stair down. The subjects wear the smart phone around the waste when performing the activities.

The results are about 90% accurate for all the other five activities except for the stair up activity and stair down activity. It was found a bit difficult for the classifier to accurately identify these two activities properly.

More than one accelerometer sensors are also used for activity recognition in [15]. Three sensors are used and positioned on different body parts to collect acceleration data. The collected data set from ten subjects is used to detect lower body movements for recognizing the activity.

Human Gait recognition techniques is proposed using Accelerometer sensor [16]. 11 subjects are used for collection of data from accelerometer sensor embedded in cell phone is place on pocket position. Data is analyzed in time domain and frequency domain for identification of human gaits for user separation based on accelerometer data.

The results were evaluated and it was found lazy learning, random forests and ensemble learning based approaches to be promising in terms of activity classification accuracy, model building time for automatic classification, and confusion matrix, with experimental validation on publicly available activity recognition dataset

The novelty of our work novelty lies in the customization of two different data sets combined for verification of the proposed framework. The framework is general for all the IoT devices. It can be extended as well for other types of data and devices. We have developed model of each specific class. i.e. sitting standing and walking of each user. These models are later used for verification and identification purpose to authenticate user and recognize activity

III. PROPOSED APPROACH

The overall architecture of the proposed approach is presented in Figure 1. The detailed description of the steps in proposed frame work is given in the following lines.

A. *IoT Data Acquisition.*

The first step of the proposed approach is data Acquisition from IoT devices i.e. sensors, actuators and cell phones etc. In this research effort, we focused on the accelerometer sensor data. There are two data sets that are customized [18,19].

First data set contains the data from four subjects 2 men and 2 women of the age range 28th year to 75 year, performing various activities i.e. sitting walking and standing. The collected data is of the accelerometer positioned on the waist.

The collected data is of the 8 hours of activities 2 hours with each subject. The second data set is collected from 15 participants performing the activities sitting, standing and walking where the accelerometer is mounted on the chest.

B. *Data preprocessing and Feature Extraction.*

As stated above that we customized two data sets [18,19]. In data preprocessing phase, a 1 second time window, with 150ms overlapping is generated in the first data set to derivate features of acceleration in axis x, y, and z.

The second data set is labeled and properly arranged with the required tri-axial attributes and user Ids. Rest of the attributes were removed and we just used the Tri-axial acceleration readings, class i.e activities and user ID as stated below.

1	User Id
2	X1
3	X2
4	X3
5	Class

User Id represent the identity of each subject to whom the data instance belongs. X1, X2 and X3 presents the tri-axial readings of accelerometer sensor. The Class attribute presents data of each activity i.e. sitting, standing and walking. In labeling phase data is labeled per its characteristic. These labeled data sets are further used by machine learning tool Weka for feature extraction.

C. *Training and Testing.*

After preprocessing and Feature extraction phase data sets are used for the training and testing purpose. Trained data sets are used for model creation. For testing purpose, real time IoT data sets are used for evaluation of the performance of models for user authentication or activity recognition task.

D. Model Creation.

For construction of models we use random forest classifier as it produces good results compared with other classification algorithm such as J48 for this data set. These models represent the standard samples or signature of each class which are used for the identification and authentication of any of the new instance of such type. These models are just like user profiles that are created once when initially data is collected. Later we can update and test it with new data sets.

E. Model Evaluation.

The real time IoT sensor data is used as test data set input for the model evaluation phase. When a data set is given for the authentication purpose its relevant model is loaded. The saved model is reevaluated on this new dataset for identification and authentication purpose. The revaluation results decide about the successful or failure of the authentication process. The overall architecture of the proposed IoT data analytics framework is presented in Figure 1.

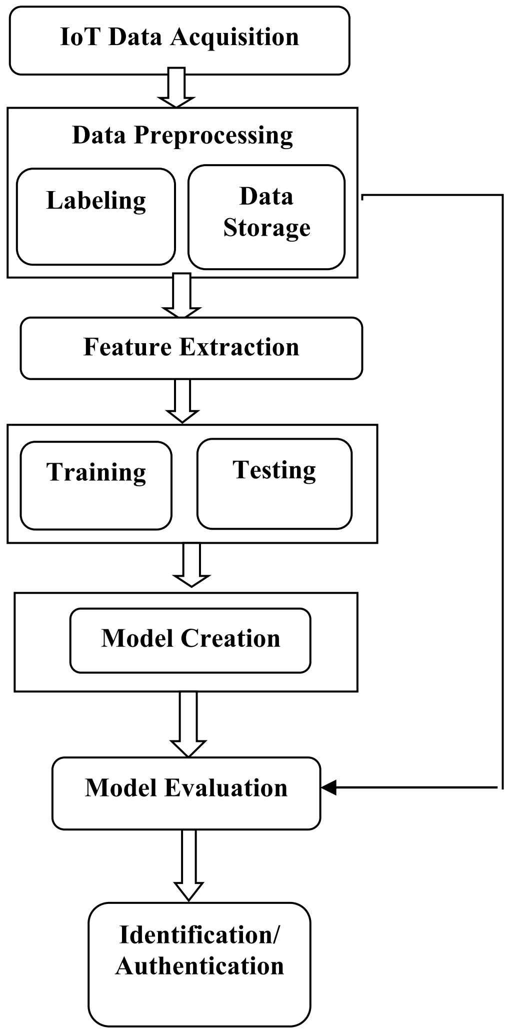

Fig 1: Architecture of the proposed Framework

IV. RESULTS AND ANALYSIS

For the prototype implementation of the proposed framework we collected the real IoT data sets [18,19]. The dataset contains the accelerometer sensor's tri-axial readings from 19 different subject performing the following three basic activities sitting, standing, and walking. We have combined the sensor readings with same attributes from two different data sets.

One data set is collected by accelerometer sensor mounted on chest and second data set is collected by accelerometer set on the waist performing the same activities. For more comprehensive analysis, different data sets collected in different scenarios for the same activities makes the experiments and results more realistic and accurate.

Another important aspect of data sets is that each activity and each user's data is balanced and with equal weightage i.e. same number of instances are used of each user and each activity. So, we can have most realistic results where variation in results represents more accuracy. We created models for each activity and each user using Weka [20].

A. Exeperiment 1.

In the first experiment, we use random forest algorithm for activities model creation. The data set include all the three activities sitting, standing and walking of 19 different subjects. Equal number of examples for each subject and each activity are presented in the data set.

Table 1 presents the results of the experiment. It applies random forest algorithm with 10-fold cross validation.

First column presents the total no of example data instances. Correctly classified instances are presented in the second column whereas the accuracy of the Activity recognition model is 93.1255%.

TABLE 1: ACTIVITY RECOGNITION MODEL USING RANDOM FOREST 10-FOLD CROSS VALIDATION.

Total No of Instances	**Correctly Identified Instances**	**Accuracy**	**Kappa Statistics**	**Mean Absolut e Error**
34999	32593	93.1255 %	0.8811	0.0587

The confusion matrix of the experiment for activity recognition model is presented in Table 2. Vertical columns represent the instances in no that are predicted by the classification results of activity recognition. Whereas in horizontal columns the actual instances are represented and compared with predicted.

In the first row, out of 3199 instances of class a.3172 are predicted accurately and 10 are predicted by the classifier as belong to b and 17 are predicted to belong to the c class.

Whereas these are a class instances. In second row, out of 15384 instances of b class 9 are predicted to belong to class a and 1196 are predicted to fall in class c rather than the actual class which is b. 14179 b class instances are accurately detected to belong to class b the actual class.

TABLE 2. CONFUSION MATRIX OF ACTIVITY RECOGNITION MODEL.

Actual class/ Predicted class	**a**	**b**	**c**
a	3172	10	17
b	9	14179	1196
c	10	1164	15242

B. Experiment 2.

The second experiment is for user identification model creation using the same data set as described in experiment 1. The user attribute is used for the classification purpose. There are 19 different subjects. Random forest algorithm is used with 10-fold cross validation. The results compared to activity recognition are less accurate with same number of instances.

TABLE 3. USER AUTHENTICATION MODEL USING RANDOM FOREST 10-FOLD CROSS VALIDATION.

Total No of Instances	**Correctly Identified Instances**	**Accuracy**	**Kappa Statistics**	**Mean Absolute Error**
34999	29869	85.3424 %	0.844	0.0194

85.34% of accuracy with kappa statistics 0.844 and mean absolute error in 0.0194

C. Experiment 3.

In this experiment user Identification model is reevaluated using the user supplied test data set. The supplied data set contains the 97 different examples of user 13.

These examples are not used for the training process while creating the model. The model is re-evaluated using the supplied test data set. Table 4 presents the results of the experiment. Table 4 presents that out of 97 examples correctly identified examples are 79. Accuracy is 81 %. With minimum number of instances used for testing as compared to training process the results are very prospective.

TABLE 4: USER AUTHENTICATION MODEL VALIDATION USING SUPPLIED DATA SET.

Total No of Instances	**Correctly Identified Instances**	**Accuracy**	**Kappa Statistics**	**Mean Absolute Error**
97	79	81.443 %	0	0.0256

Complete confusion matrix of the experiment cannot be presented here due to space constraints, 6 instances of id13 were predicted as to be Id 19,5 to Id 14,3 to Id 17. And one instance of Id 13 was predicted to belong to Id 8,9,11 and Id 16 each. The results are quite potential and if it can be further analyzed with majority component etc. it can produce 100% accuracy and can be very useful for adding an extra layer of security for user authentication.

D. Experiment 4.

For cross validation of the results and to ensure the accurate working of the algorithm we re-evaluated the user identification model with the data set of user 8 by intentionally labelling it to belong to user 11.

The algorithm worked accurately by identifying all the instances relating to user 8 rather than user 11. The confusion matrix clearly indicates that the model predicted it to be as of user 8 instances.

TABLE 5: USER AUTHENTICATION CROSS VALIDATION WITH SUPPLIED TEST DATA SET.

Total No of Instances	Correctly Identified Instances	Accuracy	Kappa Statistics	Mean Absolute Error
119	119	100%	0	0.1053

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a framework for IoT data analytics. For demonstration of the proposed framework a prototype implementation using Weka is performed on the accelerometer sensor data collected from real world subjects performing sitting, standing and walking activities. We create models of each user and each activity i.e. sitting, standing and walking, for activity recognition and user authentication. These models are reevaluated using supplied test data sets. The results obtained by the above experiments are very prospective. The results elaborate the potential use of IoT sensor data for user authentication and activity recognition.

The potential applications include healthcare, smart home, remote patient monitoring etc. It can also add an extra security layer in security sensitive systems and applications. The IoT data analytics framework is generic and in the future, work we will extend it and investigate other IoT devices and sensors for such useful applications.

ACKNOWLEDGMENT

I would like to thank my supervisor for his guidance and support throughout this research study.

REFERENCES

[1] Y.Qin et al, “When things matter: A survey on data-centric internet of things, ” J. Netw. Comput. Appl. 2016, vol .64, pp.137–153.

[2] D.Gill et al , “ Internet of Things: A Review of Surveys Based on Context Aware Intelligent Services ,” J.Sensors ,vol. 16 (17) , 2016 , pp.1-23.

[3] B.Guo et al, “Opportunistic IoT:exploring the harmonious interaction between human and the internet of things,” J.Netw.Comput.Appl 2013, vol 36(6) , pp. 1531–1539.

[4] K.Ashton., “That “Internet of Things” Thing: In the Real World Things Matter More than Ideas ,” J. RFID, 2009.

[5] G.Kecskemeti et al, “ Modeling and Simulation Chellanges in Internet of Things,” IEEE cloud computing, 2017, vol.(1) , pp.62-69.

[6] Media and Entertainment,.Internet of Things, Human Machine Interaction that unlock possibilities, Available: http://www.ey.com/gl/en/industries/media---entertainment/ey-internet-of-things.

[7] A.James et l, “ Research directions in data base archi-tectures for the internet of things: acommunication of the first international workshop on data base architecture for the internet of things (DAIT2009) ,” In: Proceedings of the 26th British national conference on databases (BNCOD). Birmingham, UK:Springer;2009.pp.225–33.

[8] A.Bayat, M.Pumplin, and D.A.Tran, “ A Study on Human Activity Recognition Using Accelerometer Data from Smartphones,” 11th International Conference on Mobile Systems and Pervasive Computing(MobiSPC) 2014, pp 450-457.

[9] A. J. Jara et al, ``Internet of Things for cultural heritage of smart cities and smart regions," in Proc. IEEE 29th Int. Conf. Adv. Inf. Netw. Appl.Workshops (WAINA), Mar. 2015, pp. 668_675.

[10] J.R.Kwapisz , G.M.Weiss and S.A.Moore, “Activity recognition using cell phone accelerometers”, ACM SigKDD Explorations Newsletter, 2010; vol.12(2) pp.74-82.

[11] A.Primo et al, “ Context-Aware Active Authentication Using Smartphone Accelerometer,” Patteren Rcognition workshop(CVRP) 2014, pp.98-105.

[12] Ferhat et al , “ Physical Human Activity Recognition Using Wearable Sensors” Sensors 2015, vol .15(12), pp. 31314–31338

[13] M.Shoaib et al, “ A Survey of Online Activity Recognition Using Mobile Phones” Sensors 2015, vol.15, pp. 2059-2085.

[14] G.Chetty, M.White and F.Akhter, “Smart Phone Based Data Mining For Human Activity Recognition,” ELSEVIER, International Conference on Information and Communication Technologies (ICICT) ,2015 , pp.1181-1187.

[15] N.C.Krishnan et al Panchanathan , “ Real time human activity recognition using tri-axial accelerometers” Sensors, signals and information processing workshop, Sedona, 2008.

[16] H.M.Thang,V.Q.Viet, N.D.Thuc and D.Chui, “ Gait identification using accelerometer on mobile phone,” International Conference on Control Automation and Information Sciences(ICCAIS), 2012

[17] W.Ugulino et al, “Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements, ” Proceedings of 21st Brazilian Symposium on Artificial Intelligence. Advances in Artificial Intelligence - SBIA 2012. In: Lecture Notes in Computer Science. , pp. 52-61.

[18] P.Casale , O.Pujol and P .Radeva , “Human activity recognition from accelerometer data using a wearable device, “ IbPRIA 11, 2014, Springer Verlag, pp.289-296.

[19] D.Anguita etb al , “Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine,” International Workshop of Ambient Assisted Living (IWAAL) 2012, Spain. , pp. 216-223.

[20] http://www.cs.waikato.ac.nz/ml/weka.