Building an Intelligent and Efficient Smart Space to Detect Human Behavior in Common Areas

Sagar Shelke, Jacob Harbour, Baris Aksanli
Department of Electrical and Computer Engineering, San Diego State University
Email:{sshelke, jharbour, baksanli}@sdsu.edu

Abstract—Smart spaces have become an integral part of our daily routines to improve quality of life for many different groups of people. The use of embedded systems to build these smart spaces, in combination with data analytics, can provide real-time information about the environment and how it interacts with the people in it. In this paper, we demonstrate how one embedded system that acquires data based on a 2-dimensional positional-grid, movement, temperature and vibration is used to build a smart and pervasive space. Data collected from these sensors is used for real time localization in conjunction with machine learning mechanisms to analyze human activities. We evaluate five machine learning algorithms, namely Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Naive Bayes and Artificial Neural Network applied on a dataset collected in our lab. Results show high classification performance for all methods giving up-to 99.95% classification accuracy. These patterns provide useful information about occupancy patterns, movement patterns, etc., which will be later used to allocate computational resources in the smart space accordingly. Furthermore, our implementation does not use any camera or microphone deployment, hence addressing potential privacy issues.

Index Terms—smart space, wireless sensor network, human behavior, data analysis, machine learning.

I. INTRODUCTION

Smart spaces have gained significant attention over the last years due to advancements in technology and ease of deployment. These spaces combine small and efficient hardware with data management mechanisms to provide solutions in various domains including health-care, wellness, education, etc. One of the important aspects of smart spaces, compared to traditional spaces, is that there is a constant interaction between humans and the surrounding environment and these interactions are captured by the deployed hardware. For some applications, such as health-care, it is imperative to analyze the human-device interaction to further understand the human behavior patterns in such environments. Previous studies tried to achieve this goal by using camera and/or microphone sources or wearable devices, e.g. [1], however, this raised very important privacy concerns. Furthermore, video and audio processing requires significant computation overhead, making them difficult to apply in applications that require real-time data analytics.

Smart spaces are usually shared by multiple people. These shared spaces commonly suffer from a lack of reliable data metrics for effective resource allocation and the ability to predict future change in resource needs. The ability to recognize human behavior plays an effective role in providing useful data. Management of facilities can use passive data gathering (no human intervention) such as this to create more informed decisions about where to effectively allocate staff, what portions of their facilities are underutilized, address security concerns, and when and where to allocate expensive resources such as power dedicated to HVAC systems [2] [3]. Another example is elderly health-care, where understanding and predicting user behavior patterns can provide timely and crucial information. Possible benefits include quality of life improvement, resource sharing improvement, etc. [4], [5].

In this paper, we demonstrate a smart space implementation that provides real-time user localization, activity detection and prediction. Different than previous studies, our system leverages efficient deployment of a variety of sensors along with a tightly coupled data collection and management strategy. Furthermore, we try to understand human behavior patterns via ambient sensors, that do not jeopardize user privacy or preferences, instead of camera, audio or wearable devices. Our smart system deploys a 2-dimensional positional grid intended to reliably detect humans. Sensors that detect temperature, motion and vibration allow us to accurately model human actions and transitions to and from the rooms in question. Our system leverages an efficient database implementation that helps manage real-time time-stamped data. The database is initially given preliminary ground-truth data, representing model human actions. Finally, we classify user activities and predict occupancy status of smart spaces using low overhead machine-learning methods, in order to obtain a real-time system. We found that Support Vector Machine classification algorithm gives the best result with 99.95% classification accuracy for activity detection task.

II. RELATED WORK

Previous research addressed some aspects of human behavior detection in a variety of ways. Ghosh et al. [6] used a grid of ultrasonic HC-SR04 sensors and machine learning to detect various activities based on the resulting data. However, this activity detection was limited, as their hardware was limited to positional sensors. In addition, we use motion detection and vibration detection to detect subtle differentiation between activities, such as sitting idle, working on a laptop, or talking on a phone. Mannini et al. [1] used a single accelerometer sensor to detect activities. While the system performance was good, use of an accelerometer on waist or ankle is not practical in areas where the occupants keep changing. Chawla et al.
[7] used machine learning methods to recognize user-specific activities. They collected data from 8 different users for 6 activities using a single wrist mounted module which sends data over bluetooth. Wrist mounting limits deployment of this module in real life as it restricts natural movements. In another work, Casale et al. [8] used a single chest mounted accelerometer sensor system to detect human behavior. To use this system in common spaces, we would need to provide each user with a data acquisition device and this might restrict their natural movements.

Modern day people commonly have smart-phones and smart-watches. These smart devices have many sensors embedded in them and many studies on human context recognition have taken advantage of this. Vaizman et al. [9] collected labeled data from 60 subjects and used it to classify daily in-the-wild context. To use this system in practice every user needs to install a smart-phone application and companion application for Pebble Smartwatch which increases dependency of the system on users. Davis et al. [10] used a similar system for Ambient Assisted Living(AAL) where users are required to wear a waist-mounted smartphone belt. Uncomfortable and inconvenient equipment compromises natural behavior. Another similar project by W. Ruan et al. [11] sought to localize individual humans within a space using RFID tags embedded within a room. This worked for a home environment in which all devices could be embedded with RFID tags, but would not be easily expandable to a shared public space where individuals would bring and use their own devices. To overcome this problem, some studies used methods that require audio/video sources to track the behavior of humans [12], [13]. However, this creates a very important privacy concern among the users of a smart space that is tracked with cameras or microphones.

Table I shows several other similar smart spaces used to model or predict specific human behaviors. As demonstrated in the table, they either rely on cumbersome wearable sensors or omit data analytics.

According to authors of [9], to promote real-life, working applications, research has to be done in natural and realistic settings. This satisfies four in-the-wild (capturing people’s authentic behavior) conditions: naturally used devices, unconstrained device placement, natural environment, and natural behavior content. Experimental setups should be capable of real life deployment so that variance in the model is at a minimum. All of the previously listed projects did their detection with varying amounts of user dependency/involvement, such as wearing sensors or installing a smart phone application. However, this is not possible in all practical applications. e.g. if a smartphone is being used to collect data, students in the library room and persons in conference rooms often keep their smartphones on a table rather than in pockets. In this case data acquired may harm a machine learning model that is being trained online. We collect a variety of data types from sensors embedded in the environment detecting position, movement, and vibration for the purposes of differentiating between a variety of activities and human behaviors, while still being able to detect various numbers of occupants.

III. OUR SMART SPACE DEPLOYMENT

A. Smart Space Goals

The primary purpose of this system was for modeling human behavior in shared rooms while following four in-the-wild conditions given in [9]. Monitoring the occupant
activity and position in shared spaces would allow other users to know about the current status/occupancy level, or provide help to the users of the system (such as elderly care). Data analytics could give users insight into future space occupancy, and provide short-term activity prediction. Occupancy of such shared spaces can vary widely based on time of day or day of the week, e.g., study rooms at a school may be occupied for much longer periods prior to students final exams, or a common area might be more crowded during lunchtime in a wellness institution. Data analytics would utilize trends normalized for season, time of day, day of week, etc. in order to give the system users useful metrics.

Privacy is an ever-present concern in sensor-based IoT systems [19]. These concerns are primarily due to the use of cameras and microphones in such systems to understand and model human behavior. In our system’s implementation, we address this issue by not using any recording devices. Instead, our system setup uses only ambient sensors and hence does not record any specific activity of a specific person. Although our system analyzes human behavior, we are not identifying the people involved with camera images or audio records.

B. Environmental Setup

Our system uses a Raspberry Pi as the base computation node. As illustrated in Figure 2, this Raspberry Pi collects data signals from various sensor inputs. The data is then transmitted to a server, where it is stored in a database. MQTT (Message Queuing Telemetry Transport) client-server protocol is used for data transfer due to its lightweight nature and ability to work even with weak Wi-Fi signals [20]. The database is accompanied by a Python script on the server that simply runs on the background in a separate thread, waiting for an MQTT message to be published by the Raspberry Pi. Data analytics algorithms then use this data for localization, activity detection and several user-end outputs. Figure 1 illustrates the real-world implementation and construction of the data collection environment, which is demonstrated in abstract by Figure 2. Sensors we have used are mainly HC-SR04 Ultrasound Range Sensors [21], Piezoelectric Sensor [22], Digital Microwave Sensor SKU: SEN0191 [23], Passive Infrared Motion Detector and Thermal sensor array MLX90621 [24]. Aim of the system is to localize the person inside and detect its activity. Ultrasound Sensors, Microwave Sensor, PIR sensor and Piezoelectric sensors are used for localization and to cycle power only through active components depending upon occupancy status. Thermal sensor is used for activity detection using machine learning algorithms.

C. Localization and Activity Detection

1) Localization: In Ambient Assisted Living (AAL) scenario, localization should be realized without instrumenting the persons to be tracked. Peter Hevesi et al. [25] used low cost IR sensors to track person within the smart environment. We have used low cost, low energy ultrasound sensors HCSR04 [21] to locate person within the environment in real time. Despite being relatively inexpensive and being somewhat imprecise, this is used in various proximity detection applications [6] [26]. These are placed in a grid as shown in Figure 4. The two dimensional positional grid made up of ultrasonic sensors has the ability to differentiate between numbers of occupants. Figure 3 shows object localization map where the rectangle represents our smart environment. We have mounted
X sensor array at different elevation than their Y counterpart to differentiate between the person standing and sitting on the chair as shown in Figure 1. To differentiate between sitting and standing person at the same position (x, y), we use blue and red colored dots. Figure 3a and 3b correlate real-time sensor data and localization result. Deflection of sensor y7 and y8 on the y-axis is occurring due to object in-front of them (Due to small spacing between two sensors, two sensors are deflected at once). Therefore, y position of dot is 230. X position is the output of these two sensors i.e. 80. Since X axis is at elevated height and sensor on this axis is deflected, person is standing and dot color will be Red.

2) Activity Detection: We divide activities into Static Activities (standing, sitting on chair, sitting on ground, laying on ground) and Dynamic Activities (move to the right, move to the left, move towards the sensor, move away from the sensor). The goal of the paper is to show the potential use of grid based environment with thermal sensor for localization, activity recognition and present results for detecting two basic activities.

a) Data Collection: Non contact thermal sensor MLX90621 have been used to collect data for activity recognition. MLX90621 has 120° horizontal field of view (FOV) and 25° vertical FOV with output in 4x16 array as shown in Figure 6. We can’t construct an image from such a low resolution output therefore privacy is not hampered. Each value in 4x16 matrix gives temperature in that area which was directly fed to the machine learning algorithms after flattening, without any feature engineering. With a single sensor placed on the middle of vertical wall as shown in Figure 5, we can capture only two activities which are standing (STAND) and sitting (SIT). Figure 8 shows heatmap of data collected for person standing and sitting on the chair. Sensor is programmed to get 4 frames/second with I2C interface connected to Arduino. We collected 111,225 labeled examples in ten days with five volunteers. Class STAND has 49384 labeled examples whereas class SIT has 61841 labeled examples. To capture more than two static activities, we need to increase vertical FOV by installing multiple sensors as shown in Figure 7 with all A, B, C equal to 25°.

b) Algorithms: We compared the performance of five machine learning algorithms using open source frameworks SciKit-Learn and TensorFlow.

1. Logistic Regression: Logistic Regression is a binary classification method which estimates probability of an instance belonging to the particular class. We used linear solver with regularization strength parameter C = 1.

2. Support Vector Machine (SVM): SVM which is a strict binary classifier like Logistic Regression, has been extensively used for human activity classification task [27],
one-vs-all or one-vs-one techniques are used to extend SVM for multiclass classification. We used SVM algorithm from Scikit-Learn with "rbf" kernel, "hinge" loss and regularization parameter C=1.

3. Decision Tree (DecTree) Decision Trees are very powerful algorithms used for classification [29]. Scikit-Learn uses CART algorithm to train Decision Trees and Gini impurity to check quality of split. CART is a greedy algorithm which looks for optimum split at top level and repeats the procedure for all levels. It selects feature k and threshold \( t_k \) such that it refines class prediction at each level.

4. Random Forest (RandFor) Ensemble methods combine predictions from several base estimators. Random forest is an ensemble method with Decision Tree as a base estimator. We used 10 Decision Tress as base estimators.

5. Naive Bayes (NaiveB) Naive Bayes is a supervised learning algorithm based on applying Bayes theorem assuming all features are independent. Naive Bayes classifiers are very fast and differ by the assumption they make about \( P(X|y) \) where \( X \) is feature and \( y \) is label. We used an algorithm with an assumption of Gaussian distribution.

6. Artificial Neural Networks (NN) NN structure is composed of several layers of nodes connected by weighted links. We used a three layer feed forward neural network written in TensorFlow. Number of inputs to the neural network are 64(4 x 16 output of MLX90621 flattened). Each hidden unit has 10 neurons. We used Batch Gradient Descent optimizer with batch size of 200. With no regularization, learning rate is set to 0.01 and activation function used is leaky relu.

### IV. RESULTS

To achieve best classification performance, Stratified K-Fold cross-validation (with \( k = 10 \)) technique was applied to each classifier. This reduced the variance of the resulting estimates because every data point was used in test set only once.

Performance metrics of the classifiers are shown in Table II. Since accuracy is not always good measure of the performance for classifiers, we also use \( F1 \) score. Each row in the confusion matrix represents actual class and each column represents predicted class as shown in Table III. Precision is the accuracy of positive predictions of the classifier. Recall(Sensitivity) is the portion of positive instance that are correctly detected by the classifier. \( F1 \) score is the harmonic mean of precision and recall, giving more weight to low values.

- **Accuracy** = \( \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \)
- **Precision** = \( \frac{T_P}{T_P + F_P} \)
- **Recall/Sensitivity** = \( \frac{T_P}{T_P + F_N} \)
- **Specificity** = \( \frac{T_N}{T_N + F_P} \)

where \( T_P \) is true positives, \( T_N \) is true negatives, \( F_P \) is false positives and \( F_N \) is false negatives.

Table II shows the performance of all models. We noticed a very high accuracy of 99.95% with Support Vector Machine (SVM) classifier whereas Naive Bayes classifier gives lowest accuracy of 91.67%. Accuracy of machine learning algorithms depends on bias and variance. Naive Bayes algorithm assumes that data distribution is Gaussian which introduces bias leading to low overall accuracy. Table III shows confusion matrix for best and worst classifier. Ensemble models in most of the cases outperform the base estimator. Single decision tree gives average accuracy of 99.34% on 10 folds whereas an ensemble of 10 decision trees slightly outperform giving overall accuracy of 99.67%. Deep feed forward neural network with three layers gives an average accuracy of 99.44% which is slightly less than best performing SVM model.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>Specificity</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>97.83</td>
<td>98.01</td>
<td>98.09</td>
<td>97.90</td>
<td>98.04</td>
</tr>
<tr>
<td>SVM</td>
<td>99.95</td>
<td>99.93</td>
<td>99.98</td>
<td>99.91</td>
<td>99.95</td>
</tr>
<tr>
<td>DecTree</td>
<td>99.51</td>
<td>99.22</td>
<td>99.39</td>
<td>99.28</td>
<td>99.30</td>
</tr>
<tr>
<td>RandFor</td>
<td>99.67</td>
<td>99.54</td>
<td>99.88</td>
<td>99.42</td>
<td>99.70</td>
</tr>
<tr>
<td>NaiveB</td>
<td>91.67</td>
<td>91.58</td>
<td>91.19</td>
<td>89.16</td>
<td>92.36</td>
</tr>
<tr>
<td>NN</td>
<td>99.44</td>
<td>99.45</td>
<td>99.55</td>
<td>99.31</td>
<td>99.49</td>
</tr>
</tbody>
</table>

**TABLE II: Performance Metrics For Algorithms in (%)**

Table III: Confusion Matrix. SVM (left) NaiveB(Right)

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th></th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SIT</td>
<td>61829</td>
<td>1294</td>
<td>STAND</td>
<td>56398</td>
</tr>
<tr>
<td>STAND</td>
<td>4249342</td>
<td></td>
<td>STAND</td>
<td>3867</td>
</tr>
</tbody>
</table>

For practical deployment of the model, train and test time of machine learning algorithms plays an important role. We evaluated train(on 90% of an entire data) and test time(on 10% of an entire data) for all models. Workstation used for benchmarking has Intel(R) Xeon(R) CPU E3-1270v5@3.60GHz processor and 8GB RAM. Figure 9 shows comparison of train and test time in seconds for machine learning algorithms. Out of all machine learning algorithms SVM takes highest time of 46.45 seconds. Computational complexity for SVM is \( O(m^2*n) \) where \( m \) (here 100,102) is number of the training examples and \( n \) (here 64) is number of features. Computational Complexity for Decision Tree is \( O(n*m*log(m)) \) having log relationship with number of examples leads to faster training. Therefore, decision tree and random forest takes 10.58 and 6.23 seconds respectively. To make predictions, each node in tree based algorithm requires checking only one feature making time complexity \( O(log_2(m)) \). Naive Bayes based model has training time complexity \( O(n*m) \) making it fastest to train but average accuracy is lowest. Training time for a neural network with three hidden layers trained in 100 epochs with 10 neurons in layer is approximately 90 seconds.

V. CONCLUSION

In this paper, we demonstrated the potential use of 2-D grid based smart space for human localization and activity detection. A two-dimensional grid of positional sensors allows it to more accurately model occupant position within a room. Furthermore, the lack of video and audio devices...
within the sensor network addresses concerns over privacy. We evaluated and compared the performance of six machine learning algorithms namely Logistic, Support Vector Machine, Artificial Neural network, Decision Tree, Random Forest and Naive Bayes in terms of average accuracy and training/test time. Experimental results reveal the superiority of SVM classifier for basic human activity classification task giving 99.95% accuracy with equally good precision and recall. Although, SVM classifier demonstrated good performance but it’s cubic dependency on number of examples make it suitable only for small or medium sized training sets. Ensemble algorithm Random Forest gives 99.67% accuracy which is slightly less than best classifier but is an ideal for large training sets as well as increased number of classes. Our future work will represent results for other static activities and dynamic activities. This smart space architecture can be applied to a variety of shared spaces, such as instructor and dynamic activities. This smart space architecture can be applied to a variety of shared spaces, such as instructor and dynamic activities. This smart space architecture can be applied to a variety of shared spaces, such as instructor and dynamic activities. This smart space architecture can be applied to a variety of shared spaces, such as instructor and dynamic activities. This smart space architecture can be applied to a variety of shared spaces, such as instructor and dynamic activities. This smart space architecture can be applied to a variety of shared spaces, such as instructor and dynamic activities.

REFERENCES


