2017 13th International Conference on Distributed Computing in Sensor Systems

COAR: Collaborative and Opportunistic Human Activity Recognition

Md Abdullah Al Hafiz Khan, Nirmalya Roy, H M Sajjad Hossain
Department of Information Systems, University of Maryland, Baltimore County
mdkhan1@umbc.edu, nroy@umbc.edu, riaj.sajjad@umbc.edu

***Abstract*—The new era of consumer devices ranging from smartphones, smartwatches, and smart jewelries augmented with our everyday activities and lifestyle help postulate human behavior, activity, gesture, social interaction, and gaming experience. Intelligently tasking and sharing the sensing, processing, storing, and computing tasks among those emerging consumer-friendly commodity devices based on their proximities, advocate the development of resource-aware collaborative and opportunistic smart living applications. Motivated by this emerging subsets of phenomenal applications, we first propose a finite-state machine (FSM) based human activity recognition framework which opportunistically exploits the relevant data sources from multiple heterogeneous devices to help infer a variety of user contexts. We depict a lightweight maximum entropy based classifier and exploit the a-priori conditional dependences among the feature sets to opportunistically select the right set of sensors with the most appropriate devices. Experimental results on real data traces demonstrate that our proposed Collaborative Opportunistic Activity Recognition, COAR framework helps infer the activities of daily living with $\approx$ 90% accuracy.**

I. Introduction

The sensors embedded in the smartphones, wearable devices, and even in the surrounding environments provide potential opportunities to concurrently share the sensing, processing, and computing power of multiple devices rationally, more elaborately *collaboratively* and *opportunistically* to infer the contexts of the users and the environments. The recent trend of miniaturization of embedded devices (smartwatch, earrings, smartshoes) and proliferation of consumer-friendly smart devices, selecting the most relevant sensors based on the user contexts, situations and availability of device resources helps design the substrate of human behavior, activity, gesture, and interaction recognition models for many smart living applications. Most of the models involving body area sensor networks are designed to transmit the streaming sensor data to a centralized node for data processing and computing, and activity learning, and inference tasks [1]. Due to the extreme mobility of humans and underpinning networking overheads, the centralized architecture is not always well suited for real-time activity recognition applications. To alleviate these existing problems, we aim to answer the following questions by proposing a *distributed* collaborative and opportunistic framework.

First, how do we collaboratively select heterogeneous sensor data streams from multiple smart devices to infer human activities at the opportune moment? ***Second***, how do we design a finite state model that can cope with the real-time activity recognition requirements while reinforcing the switching between multiple devices? ***Third***, what algorithmic innovations are required to implement and quantitatively showcase that the intelligent tasking and sharing of the activity inference process are beneficial, where the computational recourses are scarce?

To address these fundamental research questions, we propose a novel framework, **C**ollaborative **O**pportunistic **A**ctivity **R**ecognition framework, **COAR** for inferring Activities of Daily Living (ADLs) by using six smart devices - five smartphones and a smartwatch. COAR works on everyday consumer devices such as smartphones and smartwatches to ensure that our model is replicable and scalable across multiple commodity devices. COAR first recognizes micro-activity hand gestures selecting the appropriate subset of sensors and then helps posit them to discover macro-activity based on the other available smart devices' sensor data sources using FSM and finally infers activities with a light weight maximum entropy classifier. The main contributions of our work are as follows.

- We design a finite state machine model to help recognize micro-activities for intelligent sensor switching.
- We propose a lightweight maximum entropy classifier with Restricted Boltzmann Machine enabled stacked autoencoder for efficient feature discovery and establishing the relationship between heterogeneous feature distribution and the classification task.
- We conduct comprehensive experiments in an uncontrolled environment navigating the source of data streams and computational tasks between multiple smart devices and attest that our model achieves good performance for real-time activity recognition applications with minimal computational overhead.

II. Related Work

COAR builds on previous works on application of activity recognition and motivates the need for collaborative and opportunistic sensing. In this section, we compare and contrast COAR with the most relevant literature.

A. Activity Recognition

Smartphone and smartwatch has a large variety of sensors, among them accelerometer sensor makes these devices highly suitable for human activity recognition. Consequently, researchers investigated the use of smartphone and smartwatch accelerometer sensor for human activity recognition for a plethora of smart environment applications [2][3][4][5]. However, these works focus on using either smartphone or smartwatch accelerometer sensors and this affects the accuracy of detecting human activities and undermines the effective utilization of device resources. COAR on the other hand

2325-2944/17 $31.00 © 2017 IEEE
DOI 10.1109/DCOSS.2017.16

focuses on recognizing fine-grained daily living activities by opportunistically capturing hand gestures and ambulatory movements employing smartwatch and smartphones' sensors.

B. Collaborative and Opportunistic Sensing

Collaborative and opportunistic sensing in body sensor networks have been explored in various domains i.e, health-care [6], social interaction [7] etc. [8] extends InCense [9] toolkit to decide which microphone will be active when two or more microphones are present to collect similar audio information in a collaborative way. [10] introduced a iterative sensor selection process was iterative and relied on the feed-back from the user. COAR on the other hand, selects sensor source by forming boolean expressions from FSM and relies on feedback from our lightweight maximum entropy based classifier implementation. COAR exploits the selection of multi-device data sources and sensor data streams opportunistically to collaboratively infer the human activities while relying on the underpinning finite-state machine (FSM). Our FSM focuses on selecting the data sources by inferring the micro-activities. Overall, we focus on a distributed COAR framework where micro-activities are inferred based on intelligently tasked sensor nodes and the classification task is executed at the resource enriched device.

III. COAR Architecture

Our proposed COAR framework advocates a micro-activity driven collaborative and opportunistic human activity recognition model that can handle a multitude of seen and unseen activities by distributing the activity discovery and recognizing task among multiple devices. COAR boosts these capabilities by consolidating the context specific sensitivity of smart devices' sensors. Figure 1 shows our activity recognition framework. We now describe the each functional components of COAR in details.

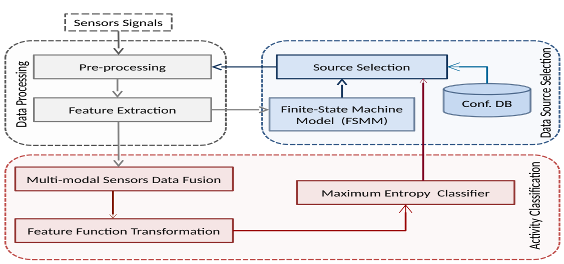

Fig. 1: *COAR Framework*

A. Data Processing

COAR data processing module has two sub-modules, data preprocessing and feature extraction.

(i) **Data Preprocessing:** This module filters raw sensor signals to remove noises using the low-pass filter (10 Hz) and provides smooth signals which help to keep activity recognition accuracy higher. The filtered data are then used to calculate corresponding features in the next step.

(ii) **Feature Extraction:** We created frames from raw sensor data in a fixed-width sliding window having a length of 3 sec per frame and 50% overlap per frame. We extracted and exploited various time domain and frequency domain features depending on the sensor types. Time domain features like mean, standard deviation etc., and frequency domain features like energy, entropy etc., of the signals, are calculated using Fast Fourier Transform (FFT) on each frame.

B. Data Source Selection

Opportunistic data source selection (ODSS) module helps to choose the appropriate sensing sources by leveraging the apriori knowledge of user activity distribution and underlying set of sensor observations. ODSS selects sensor data sources dynamically by exploiting the *configuration database* in which the corresponding mapping of the relevant data sources with respect to macro-activities are preloaded. To intelligently select the sensor data sources, we relied on a set of micro-activity states potentially help determining the potential subset of sensors from the list of available sensors. For example, when micro-activity 'sitting' is detected then the possible activities can be 'eating' or 'drinking' and these activities can be inferred using only smartwatch sensors (i.e., accelerometer, gyroscope) placed on user's wrist.

(i) **Finite State Machine Model (FSMM):** We design FSM for representing our collaborative opportunistic sensor selection strategy based on the user contexts while percolating its light-weight properties, operational efficiency and adaptability in the lack of post-processing activity information. We define the FSM model as $\overline{M} = (\Sigma, S, s_0, p, F)$ where Σ represents input feature vector from each sensor, F (set of final states) and S (a finite, non-empty set of states, such as $\{$*sitting*, *standing*, *walking*, *moving hand*, *sipping*, *stirring*, *clean table*, *washing face*, *jumping*$\}$ in our case), s_0 (the initial state which is an element of S). p represents the state-transition function: $p : S \times \Sigma \rightarrow S$, which is the core of our FSM model to select opportunistically sensor data sources. Assuming that the wearable sensor data steams follow Gaussian distribution, we define this transition probability as follows.

$$p(x_1^N, \mu_j, \Sigma_j) = \prod_{i=1}^{N} \frac{e^{-\frac{1}{2}(x_i-\mu_i)^T \Sigma_{y_i}^{-1}(x_i-\mu_i)}}{\sqrt{(2\pi)^d |\Sigma_{y_i}|}} \quad (1)$$

where there are N independent class labeled vector (x_i, y_i), $x_i \in \mathbb{R}^d$, $y_i \in \{1, 2, 3\}$, $i \in \{1, 2, ..., N\}$ and $\{\mu_j\}$ and $\{\Sigma_j\}$ represent respective class mean and covariance.

We build sensor specific FSM model for the micro-activities, we need to fuse the individual FSM model decision to infer the final micro-activity for a given testing instance. In the training phase, we emply K (K=10) fold cross-validation to computed sensitivity ($\alpha_{t,m,a}$) and specificity ($\beta_{t,m,a}$) and calculate the average sensor specific micro-activity weight as follows.

$$w_{m,a} = \frac{\alpha_{t,m,a} + \beta_{t,m,a}}{K} \quad (2)$$

For each testing instance set, we select the Maximum Likelihood (ML) represented class for each FSM. Precomputed sensor specific micro-activity weights are aggregated together from all the FSM that selects the same micro-activity. Once aggregated weights are computed, we enumerate the distinct micro-activity specific weight set for a testing instance such as $W = \{w_{a_1}, w_{a_1}, ..., w_{a_1}\}$. Next we choose the maximum weighted micro-activity from this micro-activity weighted set W.

(ii) **Data Source Selection Algorithm (DSSA):** To overcome the multiple intermittent state transitions of FSM, we consider the prior observations of the classifier and the likelihood state estimation of the FSM. We formulate the logic expressions and dynamically consolidating feedback from FSM model and classifier. Assuming that $S = \{S_1, S_2, ..., S_M\}$ is the set of sensors, the binary values 1 and 0 delineate the corresponding device which should be activated and deactivated, opportunistically. The mapping to different data sources, activities, micro-activities are also defined in the *configuration database*. This mapping is generated considering various factors such as domain knowledge, sensor modalities, sensor types, activity types, micro-activities, sensor positions etc. We devise the logic equations for each of the data sources and simplify it using boolean algebra for the final data source selection.

C. Activity Classification

In our COAR framework, we investigate a modified maximum entropy classifier (MAXENT) [11][12]. Maximum Entropy model assumes that features are conditionally dependent on each other. This is particularly useful to exploit in our case as we generate features from multiple sensor data streams across a heterogeneous set of sensors. Our micro- and macro-activity class distributions are also dependent on multiple devices sensors data streams and different chunk of data from differing devices give rise to unknown data distribution. COAR learns the conditional class distributions given the labeled training sensors data sets.

(i) **Designing Collaborative Classification:** Our goal is to construct a collaborative model that utilizes extracted features/contextual information of the activities and categorizes them to the corresponding activity classes. We assume, training set contains a large number of samples, $X = \{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$ where x represents feature vector and y is activity class label. The conditional distributions of the activity class is calculated as follows.

$$P(y|x) = \frac{1}{\sum_y \exp\left(\sum_i \lambda_i f_i(x,y)\right)} \exp\left(\sum_i \lambda_i f_i(x,y)\right) \qquad (3)$$

Here λ_i represents weight for the indicator function $f_i(x, y)$ and we formulate the indicator function follows.

$$f_{x_j,\bar{y_i}}(x_j, y_i) = \begin{cases} 1 & \text{if } \bar{y_i} = y_i \text{ and x contains code word } w_k \\ 0 & \text{Otherwise} \end{cases}$$

We assume this binary indicator feature function $f_i(x, y)$ reflects the same expected value in the training data, X so that the conditional probability $P(y|x)$ helps establish the constraints to estimate the unique class distribution and maximize the model's entropy value. The model learning parameter, λ_i is optimized by maximizing the likelihood of the training data using the exponential model as shown in Eqn. 3. This optimization is a convex optimization problem and has global maximum which we estimate using improved iterative scaling (IIS) [13] algorithm. Since in our activity recognition, $\sum_i f(x, y)$ is constant for all x and y, the partial derivative equations can be solved in closed form. We estimate the constraints from the labeled training data sets for the selected sensors.

(ii) **Multi-modal Sensors' Features Fusion:** COAR mitigates the dimensionality of variable sensors features by random projection [14] of the extracted feature vector from the selected sensor data that helps compute a fixed length feature vector. We employ feature-level fusion to increase the robustness and reliability in presence of missing sensor values which in turn helps to reduce uncertainty, and increase confidence of the performed activity. Assuming that our original signal has d-dimensions, we project it onto a k-dimensional ($k \ll d$) space using a random $k \times d$ matrix R whose column has unit lengths. The projection of the data is represented as follows. $X^{RP}_{k \times N} = R_{k \times d} X_{d \times N}$. where $X_{d \times N}$ is N d-dimensional feature vectors. We compute the R matrix using computationally efficient method proposed by Achlioptas [15].

(iii) **Formulating Feature Function:** Since, our modified maximum entropy classifier considers binary feature indicator to establish correlation between features to classify activity, we compute binary feature vector from the low-dimensional observations by employing Restricted Boltzmann Machine (RBM) [16] enabled stacked autoencoder. For the input/first layer, the gaussian visible unit activates binary activation and subsequent layers follows binary-binary RBM with Rectified Linear Units (ReLU) activation function. Autoencoder network parameters were tuned using cross-validation and the tuned settings were used in the training and testing phases in our experiment. Encoder network comprises of 4 layers which consist of two hidden layers with 1024 units each, one input layer of K units (K is determined in the fusion layer) and one output layer of 24 binary output units. Each layer is trained for 50 epochs. Since training RBM is computationally expensive, we train RBM offline and use this trained autoencoder in the testing environment.

IV. Evaluation

We now discuss the detailed evaluation of our COAR framework.

A. Experimental Setup

Classification was implemented on the smartphone. Sensor signals were sensed through our android application. 10 participants in the age group of 18 to 50 years participated in our experiment. We collected five smartphones and a smartwatch sensors data for a variety of home activities - Eating (*moving hand, sitting*), Cooking (*stirring, standing, walking*), Brushing (*walking, moving hand, washing face, standing*), Drinking (*standing, sitting, sipping*), Cleaning (*clean table, standing, walking*) and Jogging (*standing, walking, jumping (up/down)*). We captured accelerometer, gyroscope and compass sensors signals at a constant rate of 50Hz. We placed five smartphones on five different body-position - head (P_h), upper arm (P_u), waist (P_w), thigh (P_t), shin (P_s) and a smartwatch on user's dominant hand's wrist (P_a). Data from the sensors were recorded in each devices and streamed to the smartphone (P_w) placed on the waist for filtering and processing. Users performed their activities in an uncontrolled and natural environment without following any specific order or sequence. To collect the ground truth information we used skeletal tracking using Microsoft Kinect. To record activities each experiment was repeated multiple times (7 times).

B. Performance Metrics

We evaluated and compared the performance of our framework based on the following metrics. *i)* Precision =

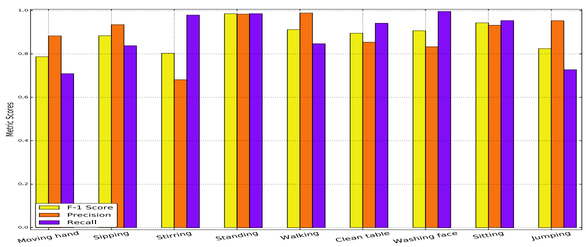

Fig. 2: *FSM metric scores for all sensors*

$(\frac{TP}{TP+FP})$, *ii)* Recall $R = (\frac{TP}{TP+FN})$, *iii)* F1 Score= $\frac{2 \times P \times R}{P+R}$ and, *iv)* Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$, where TP, FP, TN, and FN are the number of instances of true positive, false positive, true negative and false negative, respectively.

C. Inference Accuracy of FSM Model

We evaluated our FSM model based on two measures. *i)* Usage of sensor data, and *ii)* Baseline comparison.

i) Sensor Usage: We evaluated our FSM model performance in two scenarios - offline and realtime. We trained our FSM with all sensors data in the offline scenario and opportunistically selected sensors signals in realtime scenario. Fig. 2 and Fig. 3 depicts micro-activity prediction when offline collaborative and real-time opportunistic sensing strategy are employed, respectively. We note that precision for *stirring* and *washing face* are 0.68 and 0.83, respectively. This is reasonable because we have used all the sensors datasets for training and testing and our FSM model infers the maximum weighted micro-activity in decision fusion phase and detects as *moving hand*. In Fig. 3, we note that FSM model helps reduce false positives in case of selected sensors datasets.

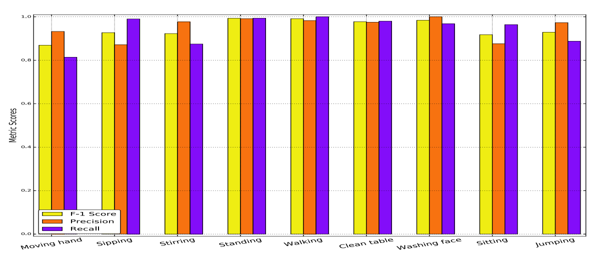

Fig. 3: *FSM metric scores for selected sensors*

ii) Baseline Comparison: We compare the performance of our FSM model with Decision Tree (DT), Random Forest (RF), Naive Bayes (NB). Fig. 4 shows comparison results on different algorithmic strategies. Our FSM model achieves 94% micro-activity recognition accuracy and outperforms NB, DT, and RF. Our FSM model infers relevant micro-activity in two phases - individual sensor based FSM, and fusion phase. In the individual sensors based FSM phase, relevant sensors datasets are used to infer micro-activities and in the fusion phase, highest weighted micro-activity is chosen among the detected micro-activities and helps achieve higher accuracy.

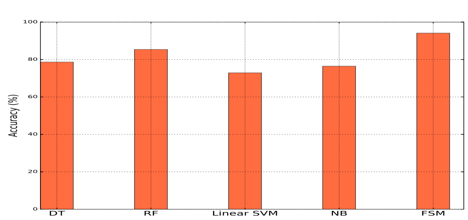

Fig. 4: *FSM performance comparison with baseline methods*

D. Classifier Inference Accuracy

We evaluated the accuracy of COAR framework using four configurations: fusing data from *i)* Standalone device (smartphone or smartwatch), *ii)* All devices (smartphones and smartwatch), *iii)* Opportunistically selected sensors devices, and *iv)* Device utilization. Case-III represents the accuracy of our model when data are fused from all the selected data entities. We fuse data samples from the sensors and pass it to the classifier for inference.

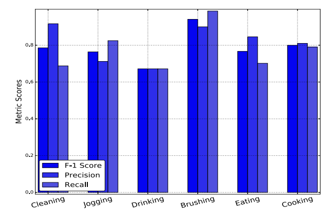

Fig. 5: *Accuracy metrics for Smartwatch*

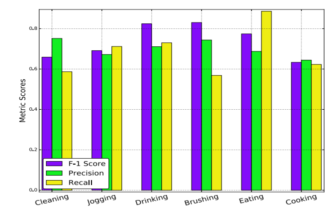

Fig. 6: *Accuracy metrics for Smartphone*

i) Standalone Device: We apply our model only on smartphone or smartwatch data respectively and compare their results as well. Fig. 5 and Fig. 6 shows average precision, recall and F1 score graph for smartwatch and smartphone sensor data. We note that cooking show poor accuracy on both devices. F1 score of jogging for the smartwatch is 0.691 whereas for the smartphone is 0.764. Though smartphone performs slightly poor than smartwatch because smartphone only captures lower extremity data well. From these observations, we can conclude that standalone devices' sensors are not sufficient to recognize activities perfectly.

ii) All Devices (Smartphones and Smartwatch): We examine our model's performance by capturing all devices sensors data at the same time and fuse these data together to feed to the classifier. Fig. 7 shows performance measurement when we employ all sensors data at the same time. Fig. 7 shows that F1 score of jogging and drinking is 0.81 and 0.80 respectively which is higher than individual sensor model. Though F1 score increases in this case but precision and recall scores show lower values. Though classifier provides high F1 score, it detects false negative instances of true jogging and drinking class instances. This motivated us to detect activity instances opportunistically and collaboratively.

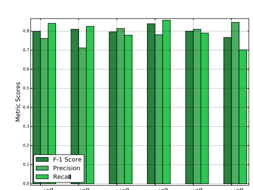

Fig. 7: *Performance metrics for data fusion*

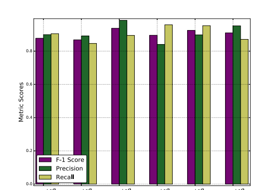

Fig. 8: *Performance metrics for real-time COAR*

iii) Opportunistically selected sensors: Our hybrid activity recognition framework opportunistically chooses sensors to infer activities. Fig. 8 shows overall performance of our COAR framework. We note that dynamic activities such as jogging and drinking have higher F1 score than using previously presented individual sensor entities. The average F1 score for jogging is 0.868 and for this case, our COAR framework employed relevant sensors from the devices. The precision scores

TABLE I: *Activity Recognition Results for Different Methods*

Activities	**DT (Accuracy %)**	**SVM (Accuracy %)**	**COAR (Accuracy %)**
Cleaning	83.40	74.78	88.47
Jogging	73.50	71.82	89.70
Drinking	84.66	75.39	88.20
Brushing	79.21	73.98	93.60
Eating	76.33	80.02	95.30
Cooking	77.23	74.78	87.00
Avg.	**79.06**	**75.20**	**90.37**

for jogging and drinking are 0.891 and 0.985, respectively, which is surely an improvement from single device inference and fusion measurement. Our collaborative and opportunistic framework achieves overall accuracy of 90.37% which is fairly good enough for a realtime system to infer activities of daily living. We compare our COAR model with Support Vector Machine (SVM), Decision Tree (DT) classifiers. Table I shows that the average accuracy for DT and SVM is 79.06% and 75.20% respectively whereas for our COAR, is 90.37%.

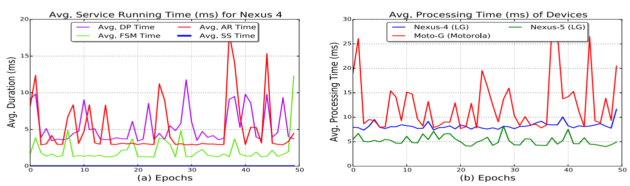

Fig. 9: *The figure illustrates time consumption of our COAR model. Figure (a) represents average time duration (per frame) of Data Processing (DP), FSMM, Activity Recognition (AR) and Sensor Selection (SS) modules. Figure (b) describes overall computational time in different devices.*

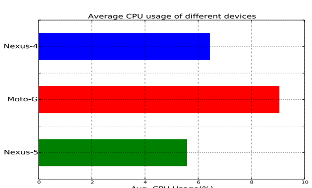

Fig. 10: *Avg CPU usage of different smartphones*

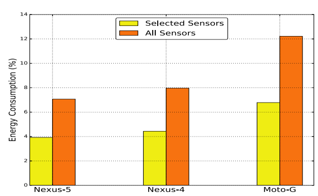

Fig. 11: *Average power consumption (%) of different smartphones*

iv) Device Utilization: We monitor the running time of our each module using Nexus-4 smartphone. Fig. 9 represents the computational time of each module and device-wise running time of our COAR framework. Fig. 9(a) illustrates average time duration of our activity recognition and data processing modules which are respectively 5.72 ms and 5.31 ms. Sensor selection (SS) module requires a constant amount of time (0.0037 ms) because of our binary sensor selection function. We also note that FSM's state detection duration (Avg. 2.32 ms) is faster compared to our overall activity recognition. Fig. 9(b) presents how fast our COAR framework performs in different smartphone models. Fig. 10 compares average CPU usage for different smartphone devices. We note that Nexus-5 requires less computational time than Nexus-4 or Moto-G as Nexus-5 has higher configuration than two of them. Fig. 11 shows average power consumption of our COAR model. We note that our COAR model consumes less energy while using selected sensors compared to using all sensors. We conclude that our activity classification process is lightweight and adaptive, and able to infer activities within a reasonable time period (5—12 ms).

V. ACKNOWLEDGMENT

This work is partially supported by the ONR under grant N00014-15-1-2229.

VI. CONCLUSIONS

In this work, we presented a collaborative opportunistic activity recognition framework using smartphone and smartwatch for emerging smart living applications. Our proposed framework opportunistically helps select sensors data from multiple heterogeneous consumer devices according to the user's current context by utilizing a finite state machine. Our data source selection module is fungible to accommodate new unseen activities that can be easily modeled using our configuration database. Our framework is flexible in adding and excluding the new and existing data sources by introducing them as new variables in the boolean expression. We validated our COAR framework in a real-life setting and our experimental results demonstrate the feasibility of collaborative sensing for real-time activity recognition applications. In future, we plan to extend our model in the presence of other consumer devices (i.e., smart necklaces etc.) and investigate the correlation between activity and other sensor modalities(i.e., audio, video etc.) to implement a light-weight version of COAR and its resource footprints for recognizing everyday activities in smart living environment.

REFERENCES

[1] Maurer and et al. Activity recognition and monitoring using multiple sensors on different body positions. In *Proc. of IEEE BSN, 2006*.

[2] Stefan Dernbach and et al. Simple and complex activity recognition through smart phones. In *IEEE IE'12*.

[3] Jennifer R Kwapisz and et al. Activity recognition using cell phone accelerometers. *ACM SigKDD'11*.

[4] Louis Atallah and et al. Real-time activity classification using ambient and wearable sensors. *IEEE Transaction in Biomedicine, 2009*.

[5] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *Pervasive computing*. 2004.

[6] Venet Osmani and et. al. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of network and computer applications*, 2008.

[7] Diane J Cook and et. al. Detection of social interaction in smart spaces. *Cybernetics and Systems: An International Journal*, 2010.

[8] Luis A. Castro and et al. Collaborative opportunistic sensing with mobile phones. In *UbiComp'14*.

[9] Castro and et. al. Incense: A research kit to facilitate behavioral data gathering from populations of mobile phone users. In *Proc. of UCAmI '11*.

[10] W. H. Wu and et al. Incremental diagnosis method for intelligent wearable sensor systems. *IEEE Trans Inf Technol Biomedicine*, 2007.

[11] Adam L Berger and et al. A maximum entropy approach to natural language processing. *Computational linguistics*, 1996.

[12] Adwait Ratnaparkhi. A simple introduction to maximum entropy models for natural language processing. *IRCS Technical Reports*, 1997.

[13] Lazebnik and et al. A maximum entropy framework for part-based texture and object recognition. In *Proc. of IEEE ICCV, 2005*.

[14] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 2003.

[15] Dimitris Achlioptas. Database-friendly random projections. In *Proc. of the ACM SIGMOD-SIGACT-SIGART*, 2001.

[16] Ruslan Salakhutdinov and et al. Restricted boltzmann machines for collaborative filtering. In *Proc. of ACM ML, 2007*.