Proceedings of 13th IEEE International Conference
on CYBER Technology in Automation, Control, and Intelligent Systems
July 11 - 14, 2023, Qinhuangdao, China

2023 IEEE 13th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER) | 979-8-3503-1519-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/CYBER59472.2023.10256442

A Q-Learning based Method for Energy-Efficient Monitoring of Activities of Daily Living using a Wearable Device and a Companion Robot

Fei Liang, Weihua Sheng

Abstract— **In this paper, we proposed a Q-learning based algorithm to trade off the activity recognition accuracy, energy cost, privacy concerns and robot resource consumption in monitoring activities of daily living (ADLs) for elderly care using a wearable device and a companion robot. The robot uses a Q-learning algorithm to decide when to turn on what sensors for data collection by considering the above four criteria. The wearable device follows the robot's decision to turn on the relevant sensors to collect data which are sent back to the robot for activity recognition. First, we described the overall design of the system in which a wearable device and a companion robot collaborate for activity monitoring. Second we developed the Q-learning framework based on the knowledge of transition motions. Third, we trained the model and evaluated the proposed method using offline data and real time data collected from our smart home testbed. The results showed that the proposed method could recognize ADLs with high accuracy while saving about half energy compared with periodic methods. The results from real time evaluation showed that the activity detection rate is about 85%.**

I. Motivation

Activities of daily living (ADLs) monitoring is critical to maintaining the independence and well-being of elderly individuals [1]. Knowing older adults' daily routine allows caregivers and family members to understand the health status and detect potential health problems of the older adults for timely intervention. Various technologies, such as internet of things (IoTs) and smart homes, have been developed for ADL monitoring. Ambient sensors [2] like video cameras and passive infrared motion sensors were employed in activity monitoring. However, it is difficult to install and maintain a monitoring system with many distributed sensors, which may pose serious privacy concerns, making them not acceptable by the older adults [3].

On the other hand, wearable devices like smart watches and smart phones are nowadays a part of people's daily life, which can be used for activity monitoring [4]. Smart watches can collect multimodal data to recognize the activities as it integrates cameras, microphones and motion sensors. For example, smart phones were used to collect heart rate and motion data for activity monitoring [5]. However, a common issue with wearable devices is energy consumption, especially when they need to collect and process large amount of data for ADL recognition. Thus, it is critical for wearable devices to collaborate with a more powerful computing resource to accomplish the recognition task. Since companion robots have been used in home care applications [6] [7], the collaboration between wearable devices and robots for ADL monitoring could achieve high recognition accuracy while reducing energy consumption.

In our previous work [8] [9], we have built a collaborative system for elderly care using a wearable device and a companion robot. The system mainly uses the wearable device to collect data and the robot to process the data for daily activities monitoring. In this paper, we consider the scenario that both the robot and the wearable device can collect data for activity recognition. In order to trade off activity recognition accuracy, energy cost, privacy concerns and robot resource consumption, the system learned a Q table to decide how to trigger the sensors given the current context. The major contributions of this paper are as follows: First, we developed a dynamic Bayesian network (DBN) based method to recognize the activities in daily life. Second, since transition motions usually indicate a change in daily activities, we treat these motions as part of the state in a reinforcement learning framework to trigger sensors so as to balance the activity recognition accuracy, energy cost, privacy protection and robot resource consumption. Third, we tested the proposed method with both offline and real-time methods in our smart home testbed.

The rest of this paper is organized as follows: Section II introduces the related work. Section III describes the overall design of the Collaborative Activity Monitoring System (CAMS) and the proposed method and algorithm for energy-efficient ADL monitoring. Section IV demonstrates the experimental setup and simulation results. Section V concludes the paper and discusses the future work.

II. Related Work

A. Activity Monitoring

A significant amount of research work has been done in the field of daily activity monitoring using various technologies such as environmental sensors, wearable devices and robots. Environmental sensors are embedded into homes to collect data related to a persons' activity, including cameras, microphones, and passive infrared motion(PIR) sensors. In the CASAS project [10], multiple sensors including infrared motion sensors, door sensors and temperature sensors were deployed in the apartments for ADL monitoring. However,

This project is supported by the National Science Foundation (NSF) Grants CISE/IIS 1910993, EHR/DUE 1928711. Fei Liang and Weihua Sheng are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, 74078, USA (e-mails: weihua.sheng@okstate.edu).

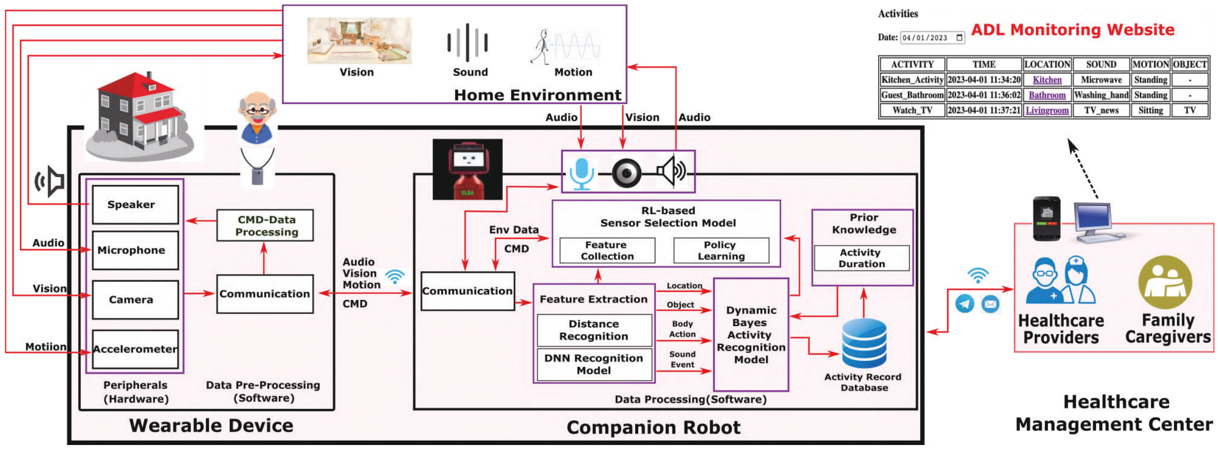

Fig. 1: The overall design of the CAMS.

the installation of distributed sensors can be costly, and maintaining the sensors can also be challenging. Nevertheless, privacy concern is the main obstacle for older adults to accept video-based or sound-based activity monitoring [3], especially when they are conducting certain activities such as using bathrooms. Companion robots and social robots are beginning to enter our daily life. With their powerful computing abilities and various onboard sensors, these robots could monitor older adults' activities and support them in their daily life [6] [11]. Georgiou *et al.* [12] developed a pet-like robot for fall detection. The robot could use its camera to recognize the activities of the person in front of it. However, using cameras again may cause privacy concerns to the user. In addition, due to the limited sensing range, the robot cannot get data if the user is far from the robot, for example, in another room. Wearable sensing and computing is becoming a promising way to monitor human activities in daily life [13]. For example, De *et al.* [14] implemented a multimodal activity recognition system by employing an accelerometer, temperature, humidity, and atmosphric pressure sensors. As a result, fusing multiple sensors could increase the recognition accuracy but at the expense of power consumption.

B. Energy Consumption Problem For Activity Monitoring

Energy consumption is a critical issue for wearable device based activity monitoring. Several projects adopted a periodic sampling method to collect data in order to save energy [15]. However, due to the periodic sampling nature, their method could miss some short-duration activities. Recently, reinforcement learning (RL) has been used for energy management [16]. For example, Possas *et al.* [17] used smart glasses for data collection and activity recognition. They utilized RL to make decisions and turn on vision or motion sensors based on the types of activities. The method was evaluated on a small dataset. In this paper, considering that motion data is highly correlated to human daily activities and could be used to classify different motion actions, we propose to use Q learning to save energy and achieve good performance in ADL monitoring.

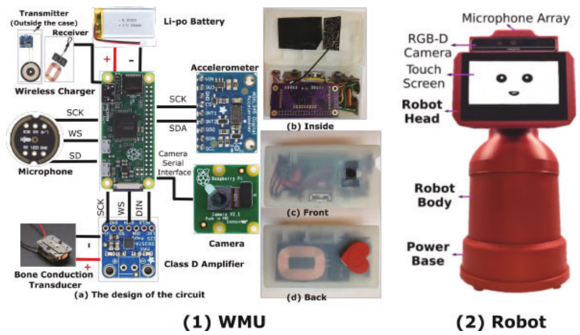

Fig. 2: The prototype of the WMU and the ASCC companion robot [18].

III. Methodology

A. Overall Design of CAMS

As shown in Fig. 1, the CAMS consists of a WMU, a companion robot, and a healthcare management system. As shown in Fig. 2, the WMU consists of three parts: the main control board based on a Raspberry Pi Zero, the battery and the housing. We integrate a camera, an accelerometer, a microphone and a speaker onto the board. With an Intel Realsense RGB-D camera and four microphones on the head, the robot can collect images and audio data for activity recognition. As shown in Fig. 1, the robot runs the RL based sensor selection algorithm to turn on the relevant sensors for data collection and activity recognition. Based on the decision, the system either turns on robot sensors or sends the commands to trigger the WMU sensors. As the WMU has limited computational resources, the data collected by the WMU will be sent to the robot for processing.

In human daily life, body motion varies when humans are conducting different activities, especially when transiting between different activities. Fig. 3 shows the changes of motion actions when conducting different activities.The transition motions would be the key to trigger the sensors and detect the corresponding activities. By default, the motion sensor is on. By monitoring the transition motion and other contextual information including current activities, battery

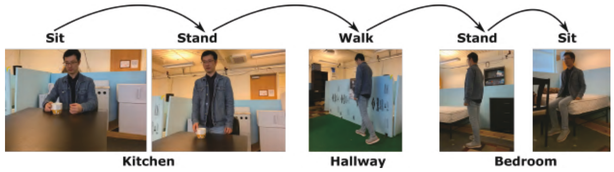

Fig. 3: The transition samples between activities.

life and robot's sensor trigger times, the agent uses the Q-Table to take actions to recognize the activities and save energy. We also consider the privacy issues when turning on sensors. For example, when in the bathroom, the WMU is a better choice to detect the activities rather than the robot. As more sensor trigger times mean more data to store on the robot, the consumption of robot resources is also considered. Then the accumulated reward is calculated from each time step. Finally, we train a Q network using an offline dataset.

B. Activity Recognition

Bayesian rule describes how to achieve the posterior probability given the evidence and the prior knowledge. In home environments, different daily activities take place at different locations, resulting in different types of sounds and body movements. As shown in Fig 4, the evidence data location L_t, object O_t, sound event S_t and body action B_t are dependent on activity A_t, based on the Bayesian rule we have,

$P(A_t|L_t, O_t, S_t, B_t) \propto P(L_t, O_t, S_t, B_t|A_t) \cdot P(A_t)$

Here, we have:

$P(A_t|L_t, O_t, S_t, B_t)$: *the posterior probability.*

$P(L_t, O_t, S_t, B_t\ |A_t)$: *the Likelihood function.*

$P(A_t)$: the prior probability of A_t.

With respect to the independent conditions, we have:

$$P(L_t, O_t, S_t, B_t|A_t) = P(L_t|A_t) \cdot P(O_t|A_t) \cdot P(S_t|A_t) \cdot P(B_t|A_t) \tag{1}$$

In this study, we estimated the duration of each activity by analyzing the dataset and combined the time label to generate the activity prior knowledge when updating the probability of A_t.

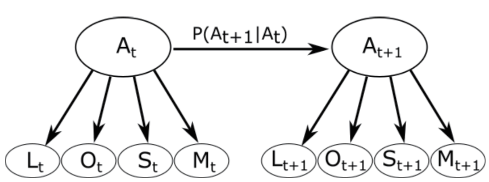

Fig. 4: The graphical representation of Dynamic Bayesian Network model for activity recognition.

Moreover, in human's daily life, the activities have sequential constraints, which can be modelled by a Dynamic Bayesian Network (DBN) model. As shown in Fig. 4, there are two parts in the DBN: the state transition model and the observation model. As shown in Fig. 5, the daily activities are modeled as hidden transition states, while the environment information - such as locations, nearby objects, sound events, and body actions recognized by the neural network models, along with the time label, are considered as observations. To recognize a target activity, we calculate the probabilities of each possible activity with a given sequence of observations, and the activity with the highest probability is the recognized activity.

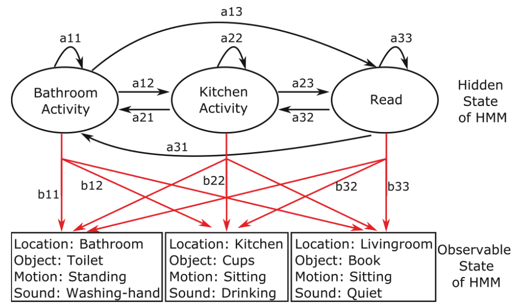

Fig. 5: An example of activities in the DBN.

C. Reinforcement Learning

Reinforcement learning (RL) is a branch of machine learning which makes decisions through trial and error by interacting with the environment. As shown in Fig. 6, in the RL process, an agent tries to learn an optimal policy which maximizes a cumulative reward by taking actions and observing the corresponding feedback from the environment. In this study, the energy consumption problem in daily activity monitoring is modelled as a Markov decision problem, in which the agent's next state depends only on the current state. The definition of state, action, reward and policy learning are as follows.

1) Goal:

- Maximize the activity detection ratio
- Minimize energy consumption
- Alleviate the privacy concerns
- Minimize robot resource consumption

2) State:

- Transition Motion
- Current Activity
- WMU Sensor Trigger Times
- Robot Sensor Trigger Times

3) Action:

- Robot Sensor: Turn on the camera and microphone on the robot
- WMU Sensor: Turn on the camera and microphone on the WMU
- Do Nothing

For the state, we have different types of common transition motions: *Walk to Walk, Walk to Stand, Stand to Stand, Stand to Walk, Stand to Sit, Sit to Sit, Sit to Stand*. We use the WMU sensor trigger times to represent the energy consumption of the battery, the robot sensor trigger times to represent the resources consumed on the robot. The agent chooses actions based on the states. For the actions, *Do Nothing* means we neither turn on the robot sensors nor the

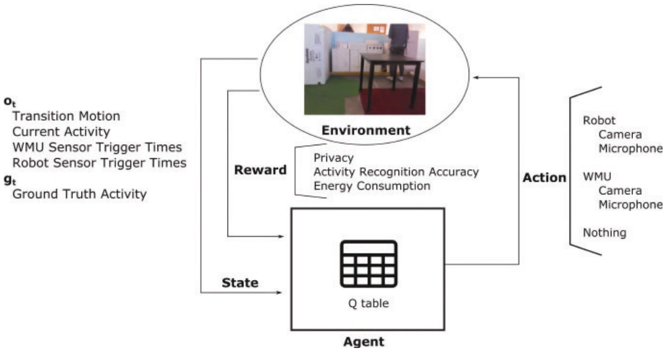

Fig. 6: The architecture of the RL based sensor selection algorithm.

WMU sensors, which saves energy during the activities. As transition motions such as *Stand to Stand, Sit to Sit* are very common, we need to figure out how to trigger the actions in order to balance the energy consumption and activity detection ratio.

4) Reward: A reward is used to estimate the benefit of a chosen action given the state. Based on the reward, the agent could learn how to select an action and get the optimal policy of the system. In this paper, accuracy of activity recognition, energy cost, privacy concerns and robot resource consumption are considered as the benefits. For each activity, accuracy of activity recognition depends on the activities and the data collected by the sensors, and energy cost depends on the amount of data collected and sent by the WMU. The resource consumption is represented by the times to turn on the sensors on the robot. For privacy concerns, we predefine some activities which could cause the privacy concerns when using the robot to capture the data. In order to learn how to use transition motions to detect the new activities, we added the factor 'extra_reward'. In summary, triggering less sensors and detecting the activity with higher accuracy will get more rewards. Table I shows the reward formula.

TABLE I: Pseudo Code for Reward Formula.

energy_cost: energy cost for data collection using the WMU sensors
resource_cost: resource cost for data collection using the robot sensors
activity_detected: 1: if activity is recognized, 0: otherwise
privacy_occur: 1: if privacy violation occurs, 0: otherwise
w_energy: the weight of energy cost and resource cost
w_accuracy: the weight of accuracy
w_privacy: the weight of privacy concern
w_resource_cost: the weight of robot resource cost
ExtraR: the extra reward when new activity is detected
1. initialization;
2. reward = -(energy_cost $\cdot w\_energy$)
+ activity_detected $\cdot w\_accuracy - privacy\_occur \cdot w\_privacy$
- resource_cost $\cdot w\_resource\_cost$;
3. **if** new activity detected **then**
4. reward = reward + ExtraR
5. **end if**

5) Policy Learning: In RL, how to learn an optimal policy that maps the state to a probability distribution of actions is important and challenging. In this paper, we employed the Q-learning algorithm [19], which is a model-free and off-policy RL algorithm. The Q table is the mapping between the states and the actions, whose values can be adjusted based on the feedback from the environment, and it uses the Bellman Equation from Dynamic Programming to optimize the collective rewards. The equation of updating Q values is as follows:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \cdot [R(s_t, a_t) + \epsilon \cdot maxQ(s_{t+1}, a_{t+1}) - Q(s_t, a_t)],$$

where Q is the action-value function with parameters state s_t and action a_t, α is the learning rate, $R(s_t, a_t)$ is the reward obtained in the state transition from s_t to s_{t+1}, ϵ is the discount factor.

IV. Experimental Evaluation

A. Evaluation of Location and Motion Action Recognition

1) Test setup: We implemented two CNN models for location recognition and motion action recognition, respectively, YOLOv3 was adopted for object recognition. To train the location and motion action recognition model, totally six locations and six motion actions are listed for classification, as shown in Table II. For motion action recognition, each sample contains 2 seconds of motion data. The proposed CNN models run on a computer with a 16-core Intel i9 CPU and an Nvidia Geforce RTX 3070 GPU, the Python and the Tensorflow version are 3.7 and 2.8.0, respectively. Fig. 7 shows the location samples from the robot view and the WMU view, respectively.

Fig. 7: The samples of locations and activities from offline dataset.

2) Results and Analysis: The overall accuracy of the location recognition model is 95% and the confusion matrix is shown in Fig. 8. We can see that the bedroom was sometimes recognized as a hallway or a door, which is due to the fact that the wall of the bedroom is similar to the wall of the hall or the door area. Similarly, the overall accuracy of the motion action recognition model is 98% and the confusion matrix is shown in Fig. 9. Specially, walking actions could be recognized correctly which is important to detecting new activities.

B. RL Model Performance Evaluation

1) Test setup: To evaluate the proposed algorithm, we conducted simulation experiments on the open smart home CASAS project [10]. Milan is selected which contains 15 activities performed over a span of 3 months in a smart home

TABLE II: The Dataset of Locations and Motion Actions

Location		Motion Actions	
Location	Samples	Motion Action	Samples
Bathroom	332	Sitting	52353
Bedroom	864	Jumping	52353
Kitchen	1207	Standing	52353
Living room	1375	Walking	52353
Hallway	280	Jogging	52353
Door Area	430	Laying	52353

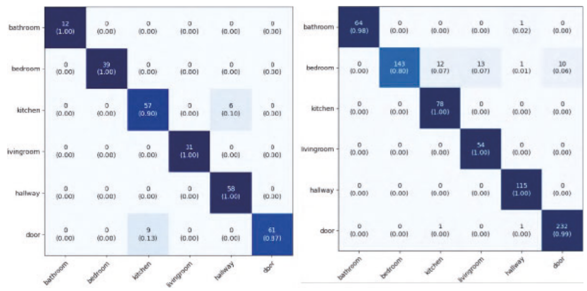

Fig. 8: The confusion matrix of location recognition. Robot view(Left), WMU view(Right).

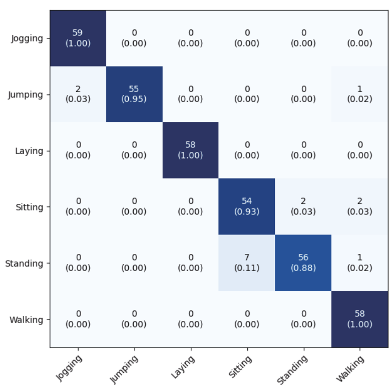

Fig. 9: The confusion matrix of motion action recognition.

where sensors are deployed at different locations. We also used the Milan dataset to calculate the prior activity probability distribution in the Bayesian network. We collected the offline dataset based on the date of 2009-12-11 and the activities are listed in Table III. The model was trained in Python3.8 on the same computer. Then, we tested the model on the data of 2009-12-11 using two different methods:

1) Periodic: The sensors are triggered periodically.

2) RL-Based: The sensors are triggered based on the predication from the RL Model.

2) Results and Analysis: The training results are shown in Fig. 10 and Fig. 11. We set the weights of accuracy, energy cost, privacy concern, and robot resource cost with 0.1, 0.4, 0.45, 0.05 respectively, which means we care more about the energy consumption and the privacy issue since the image recognition results are good with both WMU and robot data. We can see that the model converged in 3000 episodes. From Fig. 11, the trigger times of the WMU sensor converged to around 60, which is similar to the number of activities of the dataset. Also, the privacy violation occurring times converged as well.

As shown in Table. IV, for the periodic sampling method, when the trigger times decrease, the detection ratio decreases as well. Compared to a period of 2 minutes, the proposed method achieves a detection ratio of 83% with the same trigger times, which is better than the 70% achieved by the periodic method. Furthermore, to achieve a similar detection ratio of 85%, the periodic method with a 1-minute period requires the WMU sensor to be triggered 811 times. The proposed method can save 47% energy while maintaining the same level of detection accuracy.

TABLE III: Activity and Duration (in minutes).

Activity	Min	Mean	Max
Bed-to-Toilet	0.5	0.9	6.2
Chores(Vacuum Cleaning)	2.4	26.3	74.7
Desk Activity	0.5	10.8	52.8
Dining_Rm_Activity	2.35	12.2	36.7
Eve_Meds	0.2	0.5	2.1
Guest_Bathroom	0.2	2.1	16.1
Kitchen_Activity	0.2	12.3	107.2
Leave_Home	0.2	19.7	154.2
Master_Bathroom	0.2	4.9	45.1
Meditate	1.5	6.4	14.9
Watch_TV	2.1	34.3	154.3
Read	1.5	23.8	123.0
Morning_Meds	0.2	1.0	4.4
Master_Bedroom_Activity	0.2	18.6	85.2
Note: The Duration is with minute.	-	-	-

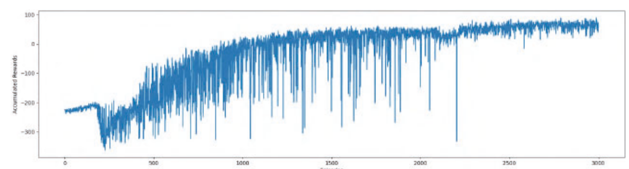

Fig. 10: Accumulated rewards.

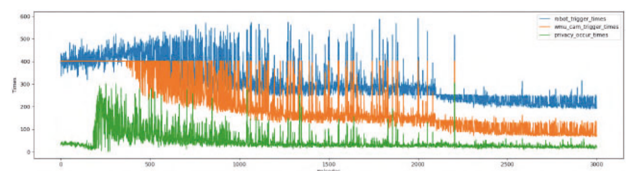

Fig. 11: Robot Sensor Trigger Times, WMU Sensor Trigger Times and Privacy Violation Occurring Times.

TABLE IV: Results between Randomly Periodic Method and the Proposed Method.

Method	Detection Ratio	Trigger Times
0.5 Mins(Period)	89%	1530
1 Mins(Period)	85%	811
2 Mins(Period)	70%	406
3 Mins(Period)	68%	271
Proposed	83%	429
Real-time	85%	-

Finally, we conducted a real-time test in the smart home testbed in our lab. Each activity lasted about 1-2 mins. The results showed that we detected 40 out of 47 activities (85% detection ration) and only missed 7 activities, which are shown as blue and yellow dots in Fig. 12. The blue dots represent the misclassified activities. When the real activity should be a 'Read' activity, the system recognized it as

'Desk_Activity'. The yellow dots show the activities we missed, which is because no data was collected. One reason is that the transition motion actions were not detected as short walk (Usually walk less than 1 second) can not be treated as a walk, while the agent generated the Q-table based on the transition motion, these activities include Read (8:50:06), Leave_Home (8:58:07), Master_Bedroom_Activity (9:00:06), Guest_bathroom (9:03:52), Guest_bathroom (9:10:38). Another reason is the poor WiFi connection, which caused the data unable to be transmitted to the server for processing. It showed time-out when transmitting data during activity Guest_bathroom (9:14:43). Generally, during the real time test, the sensors on the WMU and the robot were triggered by 67 and 95 times, respectively. During the test, we found that the transition motion types including 'stand to walk', 'walk to stand', 'stand to sit' and 'sit to stand' are helpful to detect the segmentation of the activities, the 'walk to stand' motion can indicate that human may transit to a new location, but usually after detecting the 'stand to sit', we can understand what activities was conducted.

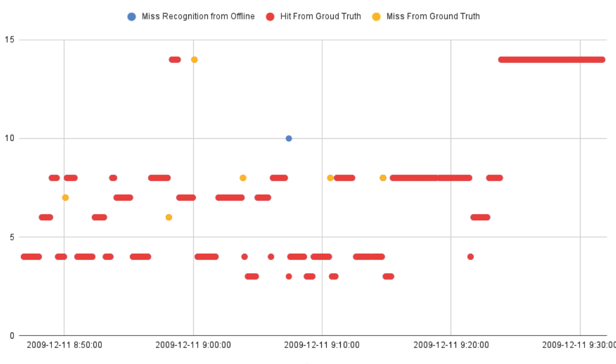

Fig. 12: The distribution of activities during one day, Watch_TV(3), Kitchen_Activity(4), Leave_Home(6), Read(7), Guest_Bathroom(8), Desk_Activity(10), Master_Bedroom_Activity(14).

V. Conclusion and Future Work

In this paper, we proposed a Q learning based algorithm to balance the recognition accuracy, energy cost, privacy concerns and robot resource consumption for ADL monitoring in a collaborative monitoring system consisting of a wearable device and a companion robot. The robot runs the Q learning algorithm to decide when to turn on what sensors to collect data. The wearable device executes the commands received from the robot to turn on the relevant sensors and send back to the robot for activity recognition. Reinforcement learning is used to solve the problem. We trained the model based on an offline dataset by reenacting the open CASAS dataset. In the offline test, the results show that the proposed method could detect 39 out of 47 while triggering the WMU sensors 429 times which saves 47% energy compared with the periodic method. In the real time evaluation, the results show that the activity detection ratio is 85%. Though this research work is promising, there are some drawbacks in this system that could be improved in the future. Particularly, we will work on these following issues: 1) Activity recognition. We will develop more advanced multi-modal activity recognition algorithms using motion, visual and audio data collected by the wearable device. 2) RL based method. We will analyze and improve the current model to include more states and actions in the RL framework.

References

[1] Activities of daily living. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK470404/

[2] M. Uddin, W. Khaksar, J. Torresen *et al.*, "Ambient sensors for elderly care and independent living: a survey," *Sensors*, vol. 18, no. 7, p. 2027, 2018.

[3] C. Berridge and T. F. Wetle, "Why older adults and their children disagree about in-home surveillance technology, sensors, and tracking," *The Gerontologist*, vol. 60, no. 5, pp. 926–934, 2020.

[4] Q. Zhou, B. Fang, J. Shan, F. Sun, and D. Guo, "A survey of the development of wearable devices," in *2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 2020, pp. 198–203.

[5] A. Mallol-Ragolta, A. Semertzidou, M. Pateraki, and B. Schuller, "harage: a novel multimodal smartwatch-based dataset for human activity recognition," in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. IEEE, 2021, pp. 01–07.

[6] I. Anghel, T. Cioara, D. Moldovan, M. Antal, C. D. Pop, I. Salomie, C. B. Pop, and V. R. Chifu, "Smart environments and social robots for age-friendly integrated care services," *International Journal of Environmental Research and Public Health*, vol. 17, no. 11, 2020.

[7] Z. Su, F. Liang, H. M. Do, A. Bishop, B. Carlson, and W. Sheng, "Conversation-based medication management system for older adults using a companion robot and cloud," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2698–2705, 2021.

[8] F. Liang, R. Hernandez, and W. Sheng, "A collaborative elderly care system using a companion robot and a wearable device," in *2022 IEEE 17th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2022, pp. 985–990.

[9] F. Liang, R. Hernandez, J. Lu, B. Ong, M. J. Moore, W. Sheng, and S. Zhang, "Collaborative fall detection using a wearable device and a companion robot," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 3684–3690.

[10] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "Casas: A smart home in a box," *Computer*, vol. 46, no. 7, pp. 62–69, 2012.

[11] A. Henschel, G. Laban, and E. S. Cross, "What makes a robot social? a review of social robots from science fiction to a home or hospital near you," *Current Robotics Reports*, pp. 1–11, 2021.

[12] T. Georgiou, K. Singh, L. Baillie, and F. Broz, "Small robots with big tasks: A proof of concept implementation using a miro for fall alert," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 206–208.

[13] S. Qiu, H. Zhao, N. Jiang, Z. Wang, L. Liu, Y. An, H. Zhao, X. Miao, R. Liu, and G. Fortino, "Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges," *Information Fusion*, vol. 80, pp. 241–265, 2022.

[14] D. De, P. Bharti, S. K. Das, and S. Chellappan, "Multimodal wearable sensing for fine-grained activity recognition in healthcare," *IEEE Internet Computing*, vol. 19, no. 5, pp. 26–35, 2015.

[15] J. Wu, Y. Feng, and P. Sun, "Sensor fusion for recognition of activities of daily living," *Sensors*, vol. 18, no. 11, p. 4029, 2018.

[16] A. Gupta and V. K. Chaurasiya, "Reinforcement learning based energy management in wireless body area network: A survey," in *2019 IEEE Conference on Information and Communication Technology*. IEEE, 2019, pp. 1–6.

[17] R. Possas, S. P. Caceres, and F. Ramos, "Egocentric activity recognition on a budget," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5967–5976.

[18] F. Erivaldo Fernandes, H. M. Do, K. Muniraju, W. Sheng, and A. J. Bishop, "Cognitive orientation assessment for older adults using social robots," in *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2017, pp. 196–201.

[19] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.