FACULTY OF INFORMATION TECHNOLOGY AND COMPUTER SCIENCES

DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

A TRANSFER LEARNING BASED TECHNIQUE FOR ACTIVITY RECOGNITION IN SMART HOMES

By:
Khaled Al-Khateeb (2015930002)

Supervisor:
Prof. Samer Samarah

THESIS SUBMITTED FOR THE FULFILLMENT OF THE MASTER DEGREE THESIS IN COMPUTER INFORMATION SYSTEMS, FACULTY OF INFORMATION TECHNOLOGY AND COMPUTER SCIENCES, YARMOUK UNIVERSITY
2018

A Transfer Learning Based Technique For Activity Recognition In Smart Homes

By

Khaled Al khateeb

B.S c. Computer Engineering/ Hashemite University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF REQUIRMENT FOR THE DEGREE OF THE MASTER OF COMPUTER INFORMAION SYSTEMS AT YARMOUK UNIVERSITY, IRBID, JORDAN.

Approved By:

Samer Samarah...Chairman

Professor of Computer Information Systems , Yarmouk University.

Mohammed Alzamil .... Member

Associate Professor of Computer Information Systems, Yarmouk University.

Mahmoud Alshboul ....Member

Associate Professor of Computer Science, Jordan University of science and Technology.

November, 2018

ACKNOWLEDGMENTS

First of all, I raise my praise and thanks to ALLAH, for supporting me with the strength and patience to start and get through this research.

I am appreciative and grateful to my supervisor Prof. Samer Samarah for his effort to guide, support and help me through this work. I am profoundly grateful to him for his valuable ideas and comments, page-by-page readings, let alone his moral support and encouragement. And many thanks to Dr. Mohammad Al Zamel for his effort and supporting me throughout this work.

I would also be thankful to my father, mother, brothers, sisters, and friends for supporting me in difficult conditions.

My greatest thanks and forwarded to my wife, who has been extremely understanding and supportive to my studies on all sides, I fell really lucky to have the best wife in the world who is willing to help me in order to reach my academic goal.

The completion of this work would not have been possible without your efforts with me.

Contents

List of Figures

List of Tables

Al-Khateeb, Khaled. A Transfer Learning Based Technique For Activity Recognition In Smart Homes, Thesis, Department of Computer Information Systems, Yarmouk University, 2018. (Supervisor:Prof. Samer Samarah).

Abstract

Human activity recognition (HAR) is a fundamental task in smart homes, that can detect and identify human activities done in smart homes, where we can apply it in many related applications like taking care of elderly people. One of the challenges of activity recognition is the need for collecting and analyzing huge amounts of data in order to be able to carry out the conventional activity discovery and recognition algorithms. This extensive initial phase of data collection and annotation results in a prolonged installation process and excessive time investment for each new space. Transfer learning addresses the problem of how to leverage previously acquired knowledge (a source domain) to improve the efficiency of learning in a new domain (the target domain) and used as a solution to handle the diversity issues for activity recognition in smart homes. In this thesis, we propose a new method for transferring learned knowledge of activities to a new domain space in order to leverage the learning process in the new environment. We applied two approaches, 1) Fixed Window Size Labeling Approach where the segmentation and labeling the data based on fixed time interval which is the average interval time for overall activities, 2) Dynamic Labeling Approach where the segmentation and labeling depend on the interval time of the detected label, so the segments length will be different depend on the detected label. To validate our algorithms, we used data collected in several smart environments with different physical layouts and different individual's behaviors. Fixed Window Size Approach shows good results when the activities have time period between them larger than the window size of segmentation. Dynamic Approach has a best result of overall accuracy for the three target domains and can deal

with various circumstances. But the two approaches could not deal well with activities that have the same nature of work.

1. Introduction

1.1 General Overview

One of the most important concerns that families facing nowadays are how to take care of their elderly parents or relatives at a stage in which they are unable of doing their daily activities or what is called Activity of Daily Living (ADLs) such as eating, dressing, sleeping, or bathing. Studies in USA show that annually 8,357,100 people receive support from long-term care services. By 2050, this number is expected to reach 27 million people; resulting in a huge demand on hospitals and health care services (Alliance, 2017).

Recently, smart homes technology has emerged as a solution to automate and control many aspects at our homes. Smart home involves the control and automation of lighting, heating, air conditioning and security. On the other hand, smart homes can help in facilitating the process of taking care of elderly residents by equipping them with sensor technologies that automatically recognize human daily activities like cooking, walking, and sleeping; allowing many applications to be developed for healthcare. Figure 1 shows a general architecture for a smart home designed to support healthcare, where the sensors read the users movements. However, machine learning algorithms allow for recognizing their activities in order to use one of the healthcare applications.

For helping the elderly people in their smart homes and using useful healthcare applications, we need first to recognize their activities in their homes through the readings of the sensors. This task called Activity Recognition. Activity recognition is a major task in healthcare- based smart homes and it aims to identifying and predicting human activities based on a series of sensor reading (Hu & Yang, 2011).It is a supervised learning problem that requires to have a set of labeled data to be used in building the model. In other words, we need to prepare data consists of records and extract features derived from the sensor readings. Each record should be attached to label; basically, one of the daily activities carried out at home. Given the diversity of smart homes architectures, labeling the data is not an easy task because of differences in sensors and differences in order to carrying out the activities.

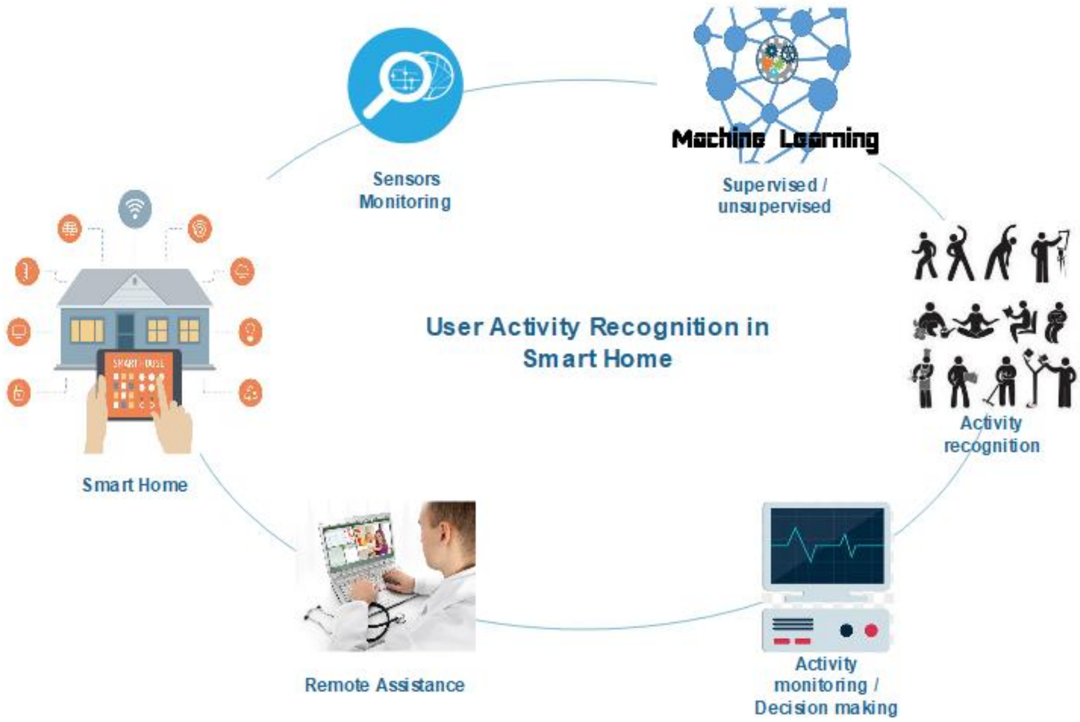

Figure 1. User Activity Recognition in Smart Home

One promising solution to deal with the shortage of labeled data is to use transfer learning. A transfer learning approaches have been proposed to explicitly deal with these sorts of circumstances. Transfer learning approach looks to apply information gained from a previous environment to a new environment. Figure (2) Shows a general picture of smart home transfer learning model, where the main tasks in the model are to initialize the features and properties of the sensors in order to apply them to the target environment. The instinct behind transfer learning stems from the ability of a human to extend what has been realized in one setting to another setting. In the field of machine learning, there are many advantages of transfer learning; less time is spent adapting new tasks, less data is required of specialists, and more circumstances can be dealt with adequately and making the educated model more robust. These potential advantages drive researchers to apply transfer learning strategies and techniques to numerous areas with varying degrees of achievement.

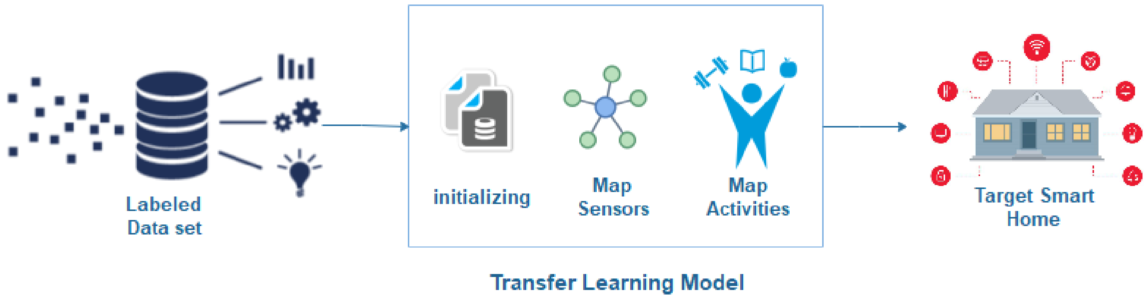

Figure 2. General Picture of Smart Home Transfer Learning

1.2 Background

The smart home is a home that is equipped with different kinds of sensors to control the home appliances and to provide monitoring for different activities at home. Healthcare is one of the applications that have been successfully implemented within smart homes; allowing to monitor and record resident daily activities. Monitoring daily activities of elderly people at home will be of a great advantage to provide the required health services and a base for alerting unhappy conditions.

Activity recognition is a major module in a healthcare system. It is the ability of the system to recognize the current activity of the resident based on the set of sensory readings. Activity recognition is a challenging task; smart homes' architectures have many diversities and people perform their activities differently. In order for an activity recognition system to successfully recognizing activities in a general design, two interesting challenges must be addressed: 1) How to adapt the system for a new environment, and 2) Resolving the diversity among different layouts and activities.

Transfer learning approaches have shown to be a promising solution to handle the diversity issues for activity recognition in smart homes, but the distributions of sensor patterns, those of activity patterns, and the maps between them are considered to be different among households. Therefore, the direct use of data from other households as training data is not necessarily very accurate (Inoue & Pan, 2016), for this reason, we need a good Transfer learning approach to address this problem, and this is our main contribution in this work which is to propose a good method for transferring knowledge within different domains that achieves good performance and accuracy.

The remaining of this thesis is organized as follows: In the next subsection we will discuss the operational definitions, chapter two defines the problem statements, chapter three summarizes related works and the major advantages and disadvantage of them, chapter four describes the proposed methodology used in our research and its implementation, the results of the proposed technique is discussed and analyzed in chapter five, and finally, chapter six summarizes the problem, our solution, and the results. Moreover, chapter six presents several ideas and suggestion to be considered in the future.

1.2.1 Operational Definitions

In this section, we will illustrate the operational definitions of the main concepts that will be used in the proposed research:

Definition 1:* *A Sensor *is a device that can detect, measure, indicate or otherwise respond to different types of inputs from environments. Set of sensors* $S = \{ s_{1,} , s_2 , \dots , s_n \}$.

Sensors could be of different forms like heat sensors, motion sensors, video cameras, thermal cameras or indoor localization systems.

Definition 2: Smart home *A smart home environment is defined as a set of sensors and activities, where activity is a recognized behavior. Set of activities* $A = \{ a_{1,}, a_2, \dots, a_n \}$.

Definition 3: Activity Recognition *is the process of assigning an activity label, from a set of predefined labels* $L = \{ l_{1,}, l_2, \dots, l_n \}$. *to a set of segmented sensors' reading.*

Sensors' readings could be reformulated based on a set of defined features. Figure 3. Shows the overall picture of activity recognition from multiple sensors.

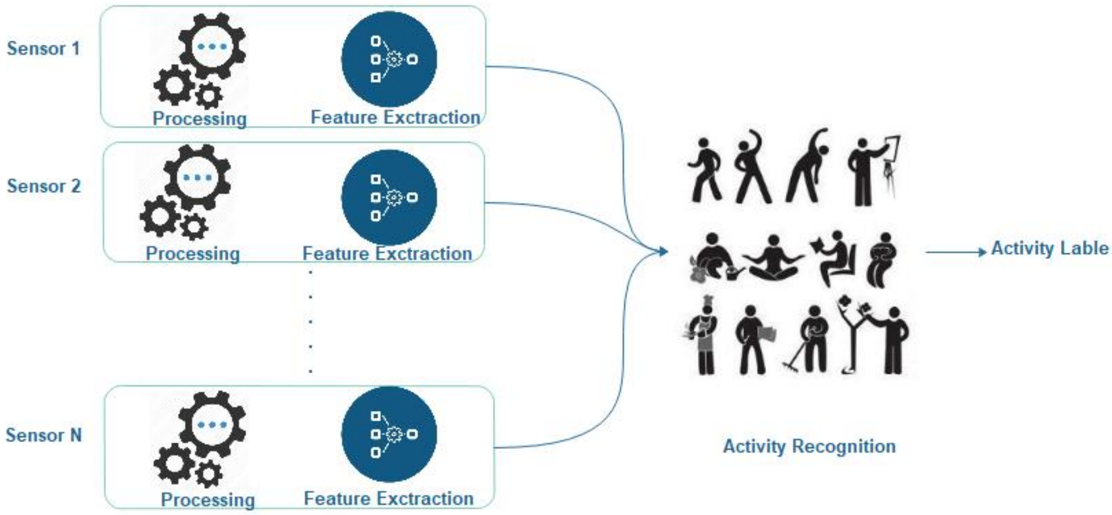

Figure 3.Overview of Activity Recognition from Multiple Sensors.

Definition 4: Activity *is a string* $\{a_1, \dots, a_i, \dots, a_n\}$ *that each element* a_i *represents a daily activity done by human like bathing, preparing breakfast...etc.*

Definition 5: Feature Extraction *start from initial measured data and build derived values which are the features, that can be used for representing the activities, it depends on the type of sensor inputs.*

Definition 6: Transfer Learning *refers to learning from one context and applying learned values to another one.*

In our research, we will use the extracted features and properties from the datasets collected using source smart home environment sensors and mapping it to the target smart home.

Definition 7: A profile *is a binary string* $\{p_1, \dots, p_i, \dots., p_n\}$ *that each element pi represents a property of a sensor or a feature in a given dataset.*

Definition 8: Data Segment *is a sequence of events that associated with a specific activity. In other words, the data segment is an activity where every sensor and event are generating a specific activity. Segmentation of data will be fixed size based on overall activity average time, or dynamically based on detected activity minimum time.*

2. Problem Statement

2.1 Research Purpose

Developing an activity recognition model in smart homes involves preparing a set of labeled data to be used in the learning phase. Given the diversity of smart homes architectures and the adhocracy nature of carrying out activities, a tremendous amount of work needed to prepare a home-specific set of training data (manually labeled data). On other hands, transfer learning shown to be a promising solution to transfer knowledge gained in one environment to another. However, transfer learning technique requires modeling every aspect in the source domain and map it into the target domain. In our research, we propose a transfer learning technique based on an abstraction of sensors' profile and define a mapping technique to relate the source and target domain. The success of defining a transfer learning technique will increase the accuracy and the performance of the activity recognition system, make it more robust and effective, and provide the reusability of the existing knowledge or training datasets.

2.2 Research Motivation

Recording the labeled datasets for each smart home is expensive and manual effort needed, which is a common problem in the field of activity recognition. Existing methods for transfer learning in human activity recognition are not good enough in terms of performance and accuracy; especially in large-scale datasets and different homes' layouts. Our goal is to solve such problems and propose a good method for transferring the training datasets and applying it to target homes with different layouts.

2.3 Research Questions

In our research, we are trying to answer the following three questions:

1- How can we transfer the knowledge in high variety domains in smart homes?

2- What is the most significant features we need to extract from the sensors readings and use them to build the features profile?

3- Is the proposed method of the transferring knowledge between the source and target domain will be better in performance and accuracy rather than manually training and labeling the target dataset?

2.4 Research Significant

Manually training and labeling the dataset of the target smart home is so expensive and very time consuming and therefore may not be the best approach either, so the main contribution of this research is to find the best way for transforming the knowledge of trained datasets of source domain to a target domain regardless to the different in the target domain i.e. different in the layout of the target domain, different in the distribution of the sensors, and the different in the individual behaviors, will be a good solution.

3. Literature Review

Human activity recognition and transfer knowledge have a significant factor in smart homes environments especially in the health care systems, in this section, we divide the related work into two subsections: First one talks about approaches used in activity recognition and its challenges. The second section talks about transfer learning methodologies and approaches used.

3.1 Activity Recognition

Human activity recognition is becoming an important field especially in the healthcare applications, and it used to provide a remote monitoring and assisting especially for elderly and Alzheimer patients who live in their own homes. The main problem in human activity recognition is how to correctly label the performed activities since we can observe the normal and abnormal activities.

Researches have proposed many approaches and methods in activity recognition problem, Hidden Markov Models (HMM), naïve Bayes Classifiers (NBC), decision tree, and support vector machine (SVM), are the common approaches used in activity recognition. According to (Mannini & Sabatini, 2010) in Markov Models, the activities are the hidden values and can be recognized from a trained data or model, where the HMM approach needs two independents assumption for tractable inference.

(Kasteren & Krose, 2007) use NBC as a classical classification model based on Bayesian theory, this approach worked well in some domains but the performance will not be good enough if there is a strong independent assumption regarding the features of the sensors used in the smart home.

Decision tree method is a commonly used algorithm for classification, it has a flowchart-like structure and its procedure is easy to understand, on another hand, it is not a useful approach to recognize a complicated activity (Parkka, et al., 2010).

About the SVM approach, the main concept is to find the best hyperplane that can be used to separate classes of high dimensional feature space, SVM with kernel function mapping can provide a good solution in activity recognition (Fleury, et al., 2010).

(Fahad, et al., 2015) proposed an approach called ARSH-SV that able to correct labeling pre-segmented activities and improve the activity recognition by reducing the incorrect assignment, while (Fang & Hu, 2014) proposed deep learning algorithm to increase the accuracy of human activity recognition.

(Rawashdeh, et al., 2017) proposed a method for activity recognition based on the activity profile by using automatic sliding windowing techniques to re-segment the stream of sensors events based in a probabilistic approach. The proposed features for building the activity profiles considered the spatial, temporal, and the semantic aspects of the source datasets. They tested different mechanisms for feature extraction and vector creation. Their results show a significant enhancement in accuracy, compared with traditional techniques. While (Hossain, et al., 2017) used active learning to enhance and reduce the labeling effort in activity recognition pipeline, they used different active learning strategies to produce a dynamic k-means clustering based active learning approach. But (Hu, et al., 2018) proposed a novel feature incremental learning method called Feature Incremental Random Forest, it consists of two components, one of them is called mutual information based diversity generation strategy to enhances the internal diversity of random forest, and the other component is feature incremental tree growing mechanism to improve the accuracy of individual decision tree. Their approach allows the dynamic exploitation of new sensors in changing environments.

3.2 Transfer Learning

Most of activity recognition approaches can introduce a good performance and accuracy if we assume that the sensors data from the source and target domains have the same distribution, however different in smart home environments and individual behaviors, or even sensors types can reduce the recognition performance and accuracy. So for reducing the time and cost effort for labeling dataset for each smart home, we need a transferring method, (Byrnes, 1996) Define transfer learning as the ability to extend what has been learned in one context to a new one. Transfer learning is studied under different names like learning to learn, knowledge transfer, and lifelong learning (Pratt, 1998).

Transferring learning can be successfully applied to independent and identically data (I.I.D) using the discriminative models which used in machine learning for modeling the dependence of target variable (y) on observed variable (x) within a probabilistic framework, and this can be done by the conditional probability distribution P(y|x) to predict the value of y from x. But in activity recognition the data are not (I.I.D) because the measurements are part of time series, and second, making discriminative models are not adequate here because we deal with large data unlabeled. Researchers have been trying to design transfer learning frameworks to identify and utilizing a good connection and transferring between the activity recognition datasets. Researchers proposed many techniques and methods for transfer learning but with some limitations in domains, performance, and accuracy. (Cao, et al., 2010) used Gaussian Mixture Model that share a prior distribution and some used the information extracted from the source house to initialize a cluster center using a k-mean algorithm (Pan & Yang, 2010). in the following sections, we summaries the transferring methods used for transfer learning.

3.2.1 Instant Transfer

Instance transfer knowledge means that we reuse the source data to train the target classifier, this usually done by reweighting the source instances based upon a given metric. (Hachiya, et al., 2012) developed an approach based on an importance weighted least-squares probabilistic classification to handle transfer learning. (Venkatesan, 2011) extended the AdaBoost framework proposed by (Freund & Schapire, 1995) to include cost-sensitive boosting which tries to weight samples from the source domain according to their relevance in the target domain. In their approach, a relevant cost given for the samples from the source domain. As the classifier is trained, those instances from the source domain with a high relevance must also be classified correctly.

3.2.2 Feature Representation Transfer

Feature representation transfer reduces the different between the source and target feature spaces. This can be accomplished by mapping the source feature space to the target feature space. In some cases, meta-features are first manually introduced into the feature space and then the feature space is automatically mapped from the source domain to the target domain (Blanke & Schiele, 2010), (Cook, 2012), (Van Kasteren, et al., 2010).

(Van Kasteren, et al., 2010) proposed three different feature mapping functions called function group to project sensors feature to a common space, after that, a semi-supervised (HMM) and improved expectation-maximization(EM) algorithms were used. Also, Kasteren used the mapping of sensors features in different houses into an individual feature space called meta feature. However, dealing with mapping the features can reduce the dimension of features and the cost of mapping computation. (Chiang & Hsu, 2012) used their background knowledge in sensors like their location, their types, and move events to assign a weight for each sensor. According to the weight of each sensor, they proposed an approach to carry out the matching between sensors without any target domain data. (Rashidi & Cook, 2010) proposed an iterative parameter updating algorithm with semi EM called home to home transfer learning (HHTL), by continuously updating the values of the matrix that describes the sensors with their activity, they can perform the transferring form source house to target house, then activity templates are constructed from the data for both the source and target data, finally, a mapping is learned between the source and target datasets based upon the similarity of activities and sensors (Rashidi & Cook, 2010). The techniques of both Rashidi and Kasteren require to define a mapping for each source and target houses, in addition to that, the manual mapping depends on the domain. (Ying, et al., 2015) proposed an approach called Multiple Cross-Domain Transfer Learning (MCDTL) for a high variety of domains by defining a metric called Kolmogorov-Smirnov for estimating the similarity of tow features if they belong to the same distribution. Based on the discovered high completeness domain set and algorithm called HCD-Miner based on a graph theory called stable matching, they could adjust the distribution of target dataset to fit the model of the source domain.

Other method used to transfer the Knowledge between different activity tasks by automatically learning a mapping of the different sensors, this done by using a Web Knowledge that provides a comprehensive citation search to help to link the different label spaces (Hu & Yang, 2011). (Chen, et al., 2017) proposed transferring learning framework based on the transformation of principles component analysis (PCA), Gale-Shapley similarity measurement, and Jensen-Shannon Divergence (JSD) feature mapping, trying to apply a new transforming method from source to target house. Their approach works will on some houses datasets but in some cases, when using only one model for all training data and obtaining the model parameters for the source and target environment and combined them with prior weight, show a better result.

3.2.3 Heterogeneous Transfer Learning

This addition transfer techniques can be applied to solve the new environments problems or the new sensing platforms problems where the source and target domain have different feature spaces. (Dillon Feuz & J. Cook, 2014) used a feature space remapping (FSR) in their approach to eliminate the need of manually mapping the features of the sensor to handle it by an algorithm to make it applicable to different domains. They used three heterogeneous transfer learning techniques: FSR, genetic algorithm for feature space remapping (Feuz & Cook, 2015), and greedy search for feature space remapping, these techniques are capable to handle different feature spaces and use a small number of labeled datasets in the target domain. Their techniques generate just a many-to-one mapping of the target dimension and not exploring ways of combining multiple dimensions. (Blitzer, et al., 2006) proposed a Structural Correspondence Learning (SCL) to use the correlation between certain pivot feature and another one to create a common presentation for these features. Also, (Daume III, 2009) transforms the features spaces of the source and the target into a high dimensional representation with components of the source, target and common, on other hand, (Diethe, et al., 2016) derive a hierarchical Bayesian model that is a natural fit the deployment context where it is likely to differ from the learning context and provide empirical validation on synthetic and publicly available datasets. And again, (Diethe, et al., 2015) proposed a framework of a combination of active and transfer learning based on hierarchal Bayesian Methods to reduce the problem of individual different in typical activities and different in house and sensor layout.

(Samarah, et al., 2018) proposed a framework that serves transferring model between different smart homes even if there is a lack of training data. They used sharing environmental characteristics in order to analyses the features between the source and target domain. Then, the features are mapped onto each other using fusion method to guarantee the variations between smart homes. They applied clustering methods for grouping related features and used the statistical correlation analysis to map the contributed features. The hidden Markov Model (HMM) has been applied to model activities in target homes. For accepting the transfer, they used a threshold value, which is an empirical parameter, they tried different values to maximize the accuracy of the overall process.

4. Methodology

Our objective is to develop a method that can transfer learned activities to the target domain, we assume that labeled activity data is available in the source domain, where the target domain contains unlabeled data. Our goal is to use the source domain knowledge to learn the activity labels in the target domain where the physical aspects and the subject's behaviors of the source and target domain are different and even if the sensors may be different also. This allows us to reduce several weeks or months of data collection and annotation in the target domain to only a few days of data collection. Our main objective is to be able to correctly recognize the activities in the target domain.

Figure 4 shows the major functions of the proposed methodology.

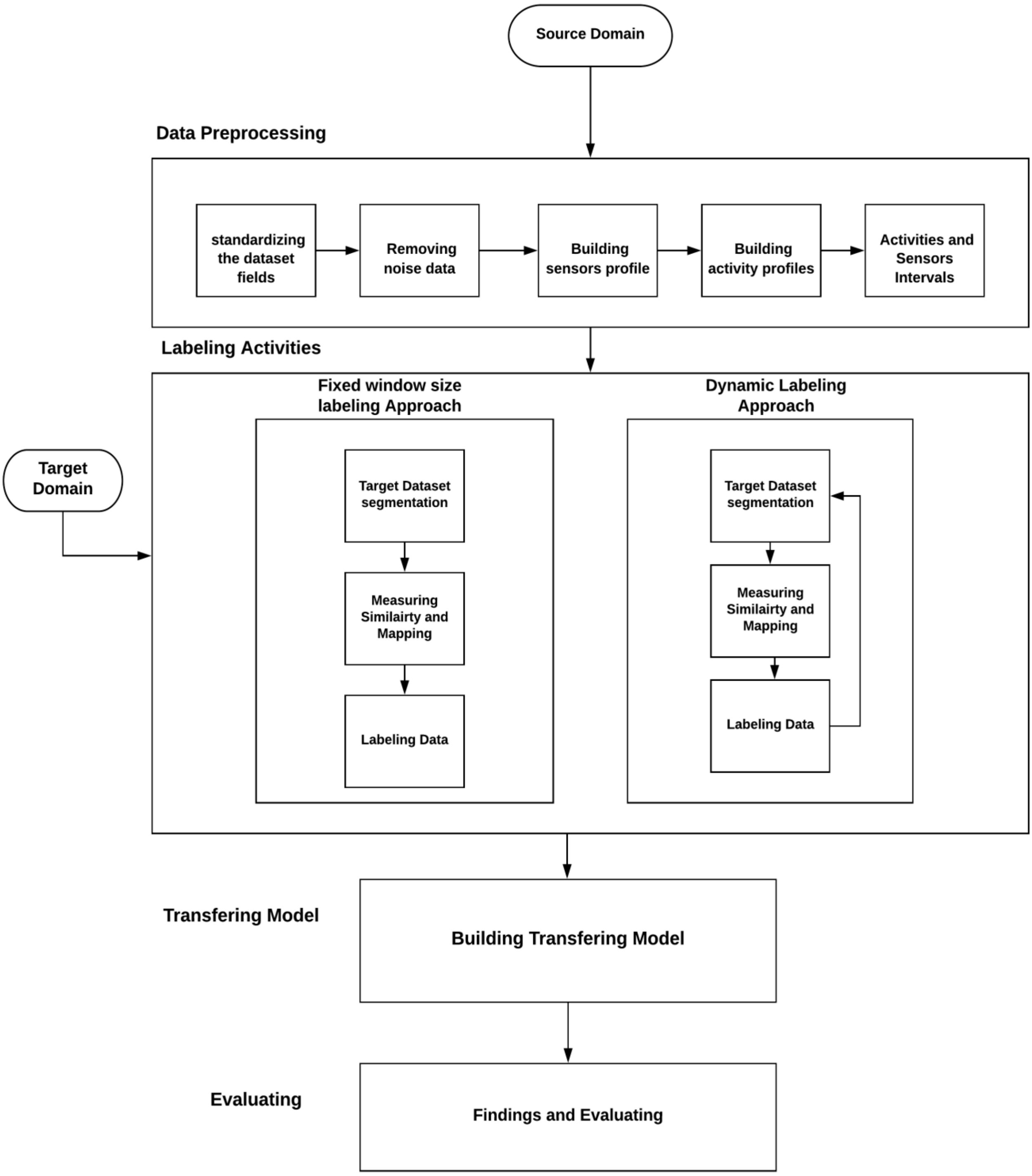

Figure 4. Overall Design of Methodology

4.1 Data Preprocessing

Our datasets contain activities for four subjects from different domains, we will use the first subject's dataset as a source domain, and the other subject as a target domain. Like any dataset, we need to do some analysis and processing on the dataset, firstly to remove the fields that are considered as a noise or faults in sensors readings, secondly, to extract features and building activities and sensors profiles to use them in our transfer model. The dataset preprocessing was carried out in the following steps:

4.1.1 Standardizing Dataset Fields

The source dataset readings contain 2772 readings with the following fields:

Table 1. Attributes Names in the Source Dataset

1	Activity label
2	date
3	Activity start time
4	Activity end time
5	Sensors ID
6	Sensor start time
7	Sensor end time
8	Sensor object
9	Sensor object ID
10	Sensor location
11	Sensor Location ID

For extracting features and removing the noise data in the dataset, we need to add some fields to the source dataset that can help us for doing these steps, so the following fields were added to the source dataset:

Table 2. Fields Added to the Source Dataset

Added field	**Description**
Activity Start Time (sec)	Convert the start time for the activities in seconds (time in 24 formats).
Activity End Time (sec)	Convert the start time for the activities in seconds (time in 24 formats).
Activity interval time	Compute the interval time for each activity in seconds.
Sensor Start Time (sec)	Convert the start time for the sensor in seconds (time in 24 formats).
Sensor End Time (sec)	Convert the start time for the sensor in seconds (time in 24 formats).
Sensor interval time	Compute the interval time for a sensor in seconds.
Year day	Convert the date of the activities into the Corresponding day number in that year.

Adding these fields in the source dataset will help us in building the sensors and the activities profiles.

4.1.2 Removing Noise Data

The second step in data preprocessing is to remove the noise data, Noise data means the readings of sensors when the sensor was kept running for a long time or ran for a very short time, and sensors that ran and not belong to any activity, like any sensor dataset from a real environment, sensors were not perfect, and so there are likely failures and noise.

In this stage for removing noise data we did the following steps:

1- Draw a box plot for every sensor, then finding the upper and lower fences, then delete the reading of sensors where their interval time below the lower fence and above the upper fence.

2- Filtering the data and finding all readings for sensors that do not belong for any activity and delete their fields.

3- Shared sensors: after filtering the sensors reading, there were 7 out of 77 sensors activated with more than one activity, the reason because that sensors keep running for a long time after the activity ends, because sensors were not perfect, and so there are likely failures.

After removing the noise data, the number of records became 2497 record.

4.1.3 Building Sensors Profile

In sensors profile, we extract the following features from the source domain to build the profile for all sensors:

Table 3. Features Extracted to Build Sensors Profile

Feature	Description
Sensor ID	ID for Each sensor
Activity Label	All the activities where each sensor activated in.
Object	The object where the sensor installed

Object 2	Another name of the object that the sensor may installed
Location	The location of the object where the sensor installed
Maximum sensor interval	The maximum length of the interval was each sensor was activated in seconds
Minimum sensor interval	Minimum length of the interval was each sensor was activated in seconds
Average sensor interval	The average length of the sensor interval in seconds

We will use the sensors profile to find the similarities between the sensors in the source and target domains.

4.1.4 Building Activity Profiles

For the source domain, we converted each contiguous sequence of sensor events with the same label to an activity profile. We have 22 different activities, so we extracted 22 different activity profiles.

4.1.4.1 Extracting Features

For each activity profile, we extracted the following features:

1- **Label:** Activity Label for each activity.

2- Sensor ID: we extracted all the sensors that activated in the same activity

3- Location: the location of objects where each activated sensor installed on.

4- Sensor Object: the object where each activated sensor installed on.

5- Occurrences Per Activity (%): the percentage of occurrence for each sensor activated in that activity.

i.e. for example, if the number of times of bathing activity in the source dataset was 18 times and the sensor ID 57 activated 10 times in bathing activity, then the % of occurrences of this sensor per activity will be 55 %.

6- Maximum Activity Interval: Maximum interval time of activity.

7- Minimum Activity Interval: Minimum interval time of activity.

8- Average Activity Interval: Average interval time of activity.

9 – Primary Object: Some activities have a primary object/objects where sensor installed on so that activity cannot be done without activating that sensor/sensors. So the value will be 1 if the sensor installed on a primary object, else, the value will be 0.

10 - Found in other activities (%): the percentage of the occurrences of the sensor in the other activities.

For example, if the sensor ID 57 found in 5 other activities out of 22 activities (because total activities are 22), so the percentage of this sensor found on other activities = 5 / 22 = 0.2272.

11 - Relevant: we created a location tag for each activity, so we choose the possible locations where each activity may happen. The following table shows the location tag for each activity.

So, the value of this field will be 1 if the location of the sensor refers to the activity regarding according to the above table, otherwise, its value is 0.

Table 4. Relevant Location(s) for each Activity

Activity	**Location Tag**
Bathing	Bathroom, Bedroom
Cleaning	Bathroom, Bedroom, Kitchen, Living room
Doing Laundry	Kitchen, Bathroom
Dressing	Bathroom, Bedroom

Going out shopping	Foyer, Kitchen
Going out to work	Foyer
Going out for entertainment	Foyer
Grooming	Bathroom, Bedroom
Lawn work	Kitchen
Preparing lunch	Kitchen
Preparing snack	Kitchen
Preparing breakfast	Kitchen
Preparing dinner	Kitchen
Preparing beverage	Kitchen
Putting away laundry	Bedroom
Putting away dishes	Kitchen
Putting away groceries	Kitchen, Bathroom
Toileting	Bathroom
Washing dishes	Kitchen
Washing hand	Bathroom
Watching TV	Living room
Other	Bathroom, Bedroom, Kitchen, Living room, foyer

12 - Weight value: to give a weight value for each sensor, we use this equation:

Weight value= Occurrences per activity (%) - Found in other activities (%).

Regarding the above equation, if a sensor found in all activity record and not found in other activities, so the maximum value will be (1), on another hand, if a sensor not found in activity records and found in all other activities, so the minimum value will be

(-1). So, we use this formula to normalize the weight value to become between **0** and **1**.

$$\boldsymbol{Xnew} = \frac{(X+1)}{2} \quad \dots\dots\dots\dots\dots\dots\dots\dots \textbf{(1)}$$

4.2.4.2 Filtering Sensors

For each activity, there are some of the sensors activated in that activity but it may be not relevant to that activity, so to filter the most relevant sensor that refer to such activity, we choose the sensors that their locations refer to that activity or the sensors where their weight value above the midpoint which is **(0.5).**

4.1.5 Activities and Sensors Intervals

In our approach we use the Time as one of the features that may help us for distinguishing between activities in detecting Labels for activities, we choose to divide the day into 8 intervals, each interval with 3 hours, starting from the midnight. Then for each activity, we find the intervals where that activity was activated. So, if an activity-activated in interval 1, the value will be 1, else, 0. And so on for all intervals and activities.

The same method was used for building sensors intervals. This can help us during finding the similarity between the source and target domain, and also reducing the number of sensors, and searched activities during the stage detecting labels.

4.2 Fixed Window Size Labeling Approach

4.2.1 Data Segmentation

In fixed window size labeling we divided the target dataset into segments based on the average time of overall the activities in the source domain which is 1123 seconds. Figure (5) shows the flow of segmentation target dataset.

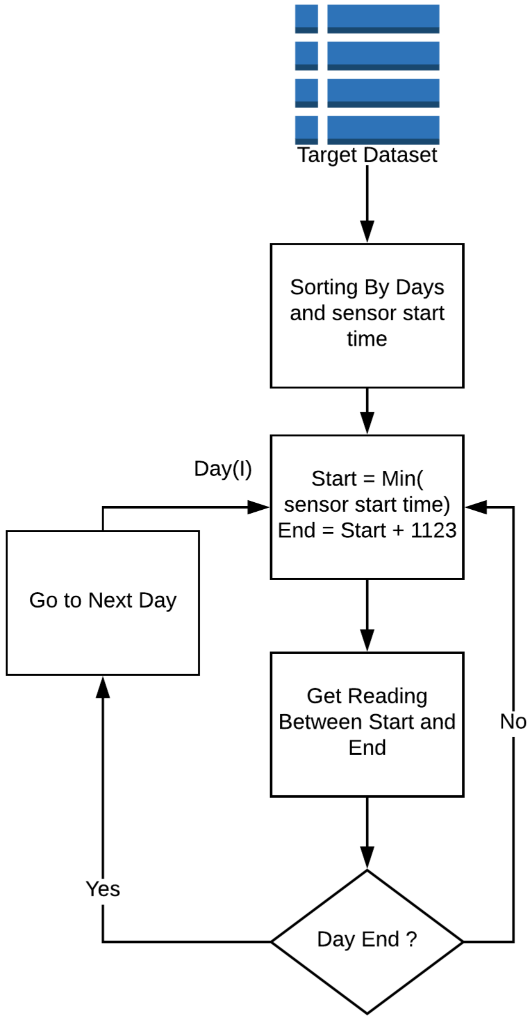

Figure 5. Flow of Segmentation of Target Dataset

For each day in the target dataset, we first sort the reading of sensors start time in an ascending order, then each segment length will equal or less than 1123 seconds, then continue to the next sensor start time. This segmentation continues until the day ends, we repeated this operation for all days in the target dataset. Each segment will be like an activity.

4.2.2 Labeling Segments.

For each segment, we tried to find the best label for it. Our approach based on finding the similarity between sensors in the segment, and the sensors in the sensor profile, then finding the best-fit label for each segment. This approach was divided into three stages:

1- Finding Similarity Between Sensors:

The similarity between sensors in each segment and sensors in the sensors profile shown in the figure (6).

We first find the similar sensors between segment and sensor profile based on the similarity in the locations and the objects of each sensor. After that we took the first sensor start time in the segment and find all the sensors that worked in that interval based on the sensors intervals, the same thing was done for the last sensor start time in the segment. After that, we find the common sensors between the three lists.

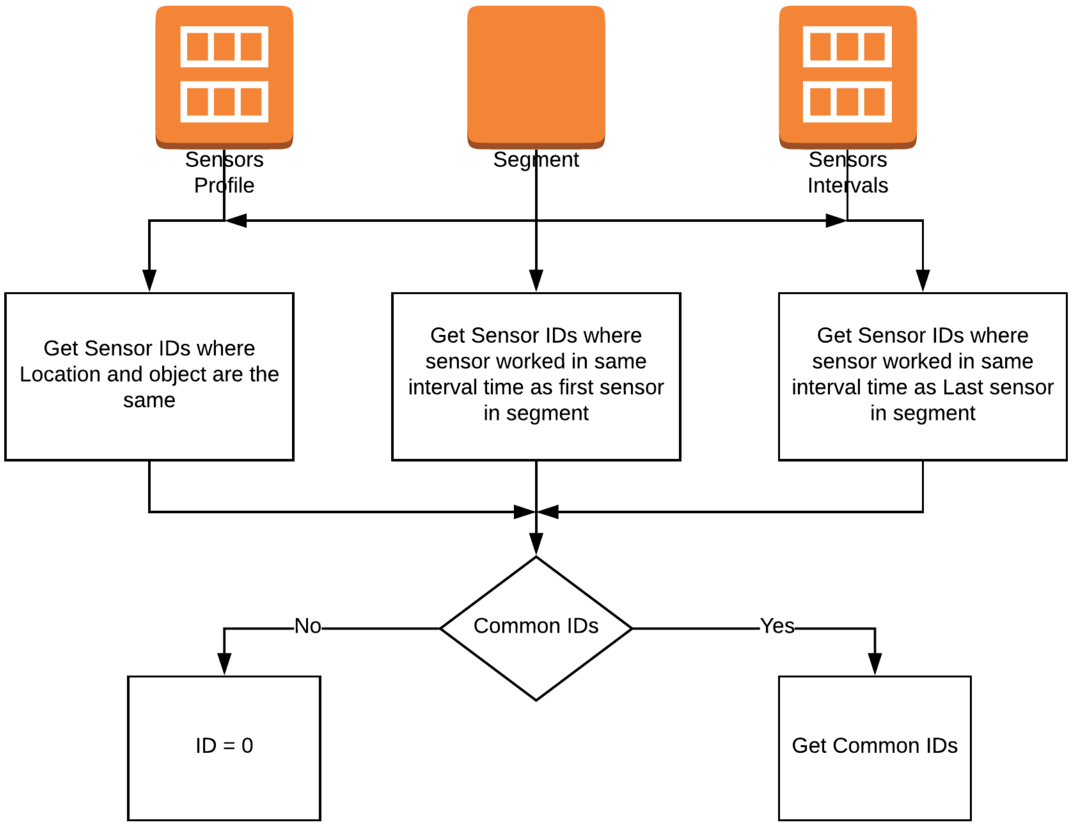

Figure 6. Finding Similarity in Sensors Between the Source and Target Datasets

2- Finding Activities Available to Search:

In this step, we trying to find the most relevant activities for each segment, then trying to find the best-fit activity for that segment. The following figure shows the way of finding the most relevant activities:

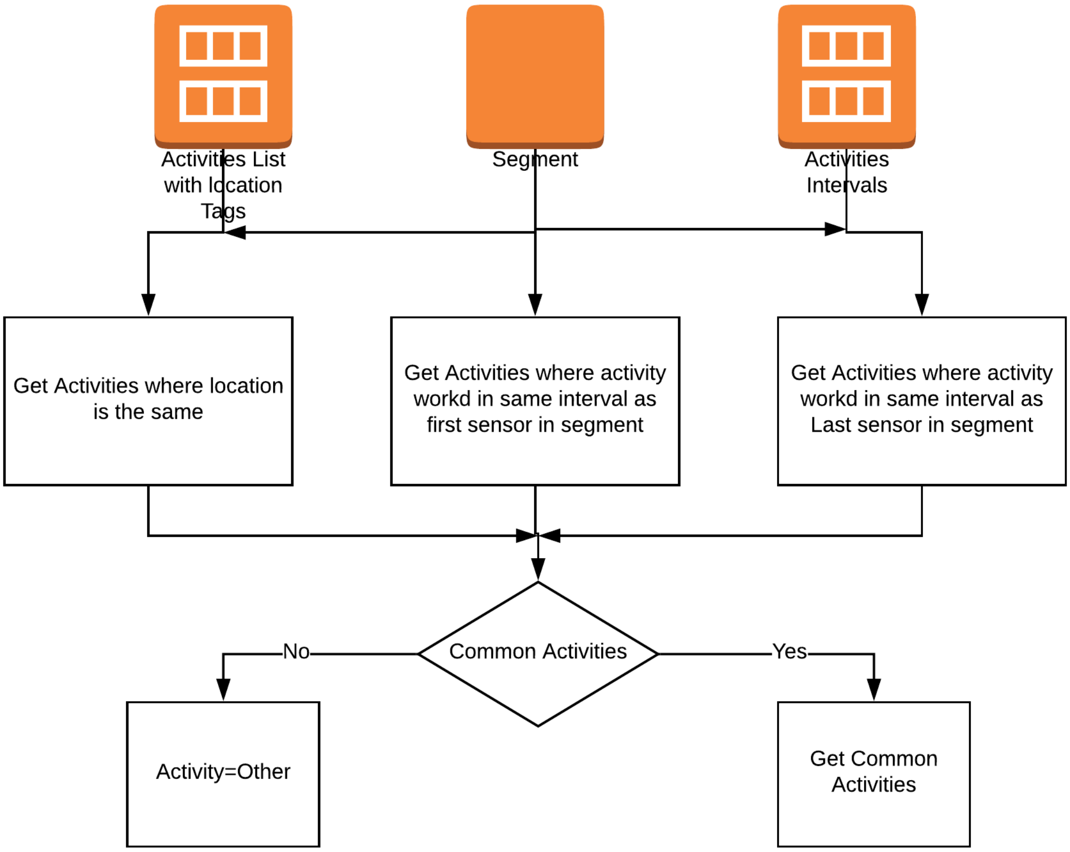

Figure 7. Way of finding the Most Relevant Activities

For each segment, we took each sensor's location and find all possible activities that may happen in that location, after that for the first sensor time and last sensor time in the segment, we find all the activities that happened in these intervals based on activities intervals. After finding all the possible activates, we find the common activities between these lists to find the best activity fits that segment. If there was no common sensors or activities between lists, the label of that segment will be “OTHER”.

3- Find Best Activity Label:

After finding the similar sensors between the segment and sensors profile, and after extracting the most relevant activities that may fit that segment, we found the best label fit that sensor based on the following equation:

$$\boldsymbol{Activity(i)} = \sum_{j=1}^{n} (\boldsymbol{a_j} + \boldsymbol{b_j}) \text{ .... } \mathbf{(2)}$$

Where: n: Number of total most relevant activities.

a_j : Occurrences per activity for sensor j, (j∈ $Activity(i)$)

b_j: the value of the primary object of sensor j , (j∈ $Activity(i)$)

So the best activity that fit the segment will be the activity with the max value.

4.3 Dynamic Labeling Approach

In the dynamic labeling approach, the segmentation and labeling the data will not base on fixed size of activity length, but choosing and moving to the next segment of data of the target dataset will be based on the Minimum Activity interval time of the detected activity.

In our source domain dataset, the minimum activity interval time was for toileting activity with (25) seconds. Figure (8) shows a general overview of the dynamic labeling approach.

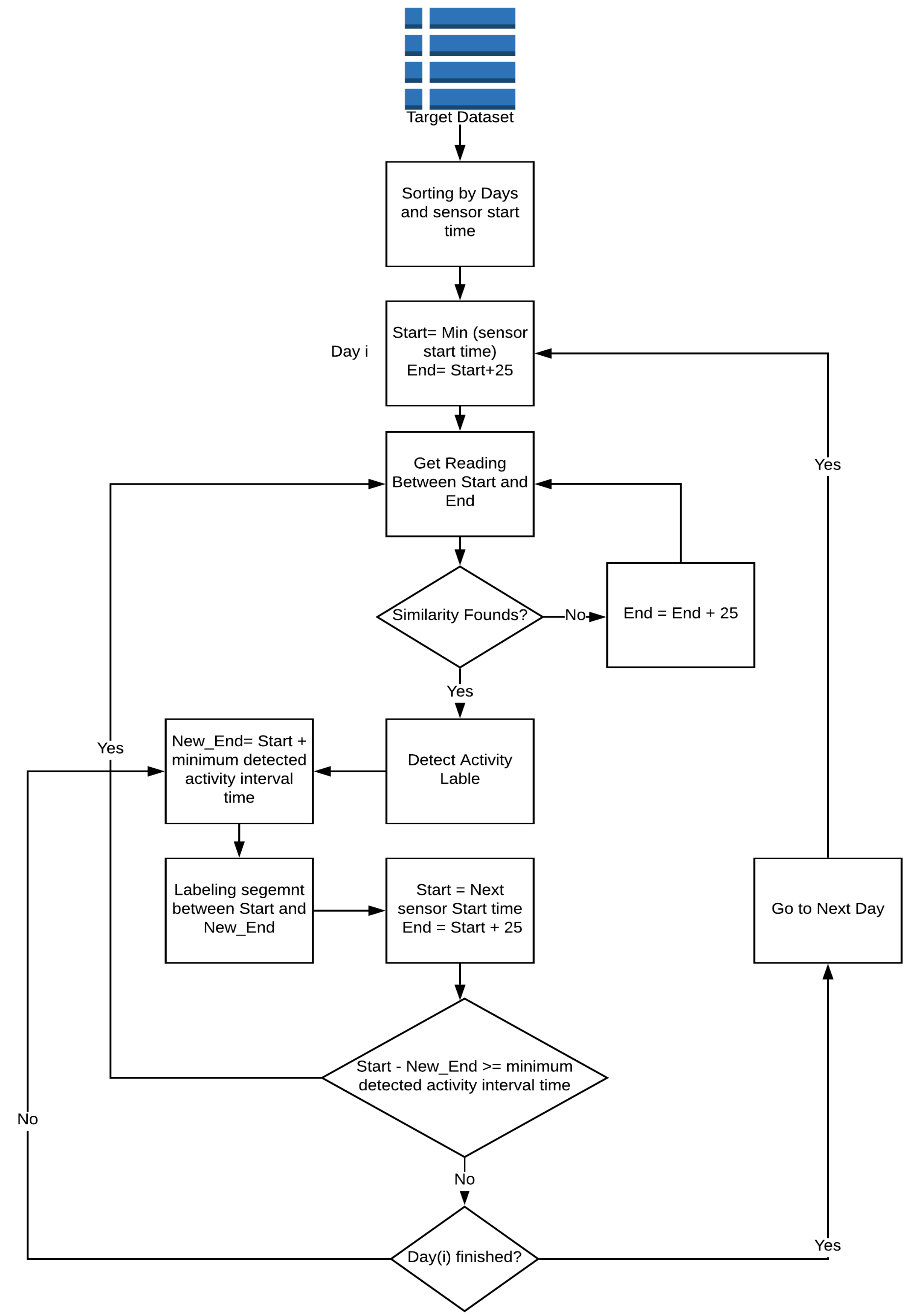

Figure 8. Dynamic Labeling Approach Design

In the dynamic approach and after sorting the dataset readings based on days and sensor's start time, the similarity between sensors and detecting the most relevant activities will be the same as in the fixed window size approach, but the different here is, when starting labeling the sensors reading, the first segment length will be the length of the activity which has minimum interval time which is (25) seconds, then finding the best label for that segment as in fixed window size approach, after that, we increase the segment interval based on the minimum interval time of the detected label, we will keep increasing the segment interval until the different between the value of the sensor start time of the next segment and the shift value become greater than or equal the minimum interval time of the detected label, then we move to the next segment. In another word, we move to the next segment if the sensor different between sensor start time and lower limit of the segment is larger than or equal to the double value of minimum interval time.

This condition will decrease the possibility of overlapping between activities. After detecting the label for the first segment, we continue to the next sensor's start time for the next segment with length equal to the length of the activity which has minimum interval time and continue the iteration as before. The loop will continue until the readings of the first day ends, then continue to the next day and so on.

If there is no common sensors or activities in the lists during finding the similarity between sensors or during finding the most relevant activity for each segment, we double the length of the segment until getting common sensors or relevant activities.

This approach will increase the percentage of similarity because the overlapping between activities will be decreased, as we will see in the experiment and result section.

The following example shows the approach in more details.

For example, let say that we have the following sensor start time sequence:

Table 5. Sample of Sensor Start Time Sequence

Sensor ID	Sensor Start Time
141	27596
141	27640
112	27807
141	28594
141	28598
98	28704
115	28944
96	29955

The start time in the segment will be 27596, the first segment length will be less than or equal 25 seconds. So we get all sensors between 27596 – 27621, which is sensor 141, then find the similarity and the label of that segment, for example the detected label of the segment is "Watching TV", from activities profiles, the minimum interval time of Watching TV is 303 seconds, so we extend the length of that segment to become less than or equal 303, the lower limit of the segment is 27596+303 = 27899. So the new segment will start from 27596 to lower limit which is 27899, the label of all sensors between this interval will be "Watching TV" as shown in the next table.

Table 6. First Segment in the Sample with its Label

Sensor ID	**Sensor Start Time**	**Detected Label**	**Segment**
141	27596	**Watching TV**	**Segment 1**
141	27640	**Watching TV**	**Segment 1**
112	27807	**Watching TV**	**Segment 1**
141	28594		
141	28598		
98	28704		
115	28944		
96	29955		

After that we move to the next sensor start time which is 28594, the different sensor starts time and the segment lower limit = 28594 – 27899 = 695 which is larger than the double value of minimum activity interval time, so this sensor start time will be the start of the next segment.

Next segment starts from 28594 to 28619 (28594 + 25). Sensors in that segment will be 141, and after finding the similarity and label, the detected label will be "Watching TV". Lower limit of segment 2 will be 28594 + 303 = 28897. So all sensors with this segment interval will be "Watching TV" label as shown in the next table.

Table 7. Second Segment in the Sample with Label

Sensor ID	**Sensor Start Time**	**Detected Label**	**Segment**
141	27596	Watching TV	Segment 1

141	27640	Watching TV	Segment 1
112	27807	Watching TV	Segment 1
141	**28594**	**Watching TV**	**Segment 2**
141	**28598**	**Watching TV**	**Segment 2**
98	**28704**	**Watching TV**	**Segment 2**
115	28944		
96	29955		

After that we move to the next sensor start time which is 28944, the different sensor starts time and the segment lower limit = 28944 – 28897 = 74 which is lower than 606 (double value of the minimum activity time of Watching TV), so we still in segment 2. The new lower limit of segment 2 will 28944+303 = 29247. So, all sensors which have sensor start time less than or equal 29247 will belong to segment 2 as shown in the next table.

Table 8. Extending Segment No.2 in the Sample

Sensor ID	Sensor Start Time	Detected Label	Segment
141	27596	Watching TV	Segment 1
141	27640	Watching TV	Segment 1
112	27807	Watching TV	Segment 1
141	28594	Watching TV	Segment 2
141	28598	Watching TV	Segment 2
98	28704	Watching TV	Segment 2
115	28944	**Watching TV**	**Segment 2**
96	29955		

After that we move to the next sensor start time which is 29955, the different sensor start time and the segment lower limit = 29955 – 29247 = 708 which is larger than the double value of minimum activity interval time, so this sensor start time will be the start of the next segment.

5. Experiment and Results

5.1 Dataset

We will use four datasets for the evaluation, the datasets were obtained using similar sensor systems in different houses. The layout of the different home settings differs strongly, as well as the sensors configuration, and the individuals who perform the activities are all different. More information about these datasets can be found in (Munguia Tapia, 2003), (Kasteren, et al., 2011), (Ordonez, et al., 2013). An overview of the datasets can be found in Table 9.

Table 9. Datasets Description.

Dataset	MAS S1	MAS S2	OrdonezB	KasterenA
No. of Sensors	77	70	12	14
No. of Activities	22	24	16	10
Setting	Apartment	Apartment	House	Apartment
Duration	16 Days	17 Days	21 Days	22 Days
Rooms	4	5	5	3

We will use the dataset MAS S1 as the source domain the other datasets as the target domains. Target domains activities are already labeled but will hide these labels and assume that the target domains are unlabeled. We will do the data preprocessing and features extraction on the source domain, and then we will apply our approach to the target domains and find the percentage of the Total Accuracy between original labels and detected labels.

5.2 Data Preprocessing

The original source domain contains about 2772 records, but before extracting features we need first to standardizing the data and removing noise data.

5.2.1 Standardizing Source Domain

We need to add some fields to the source dataset that can help us for extracting Features. The source domain will have the following fields: *SDomain= {Activity Label, Activity StartTime, Activity StartTime(sec), Activity EndTime, Activity EndTime(sec), Activity IntervalTime, Date, Year Day, Sensor ID, Location, Location ID,Sensor Object, Sensor Object ID , Sensor StartTime, Sensor StartTime(sec), Sensor EndTime, Sensor EndTime(sec), Sensor IntervalTime}*

5.2.2 Removing Noise Data

Source domain contains noise data, where some sensors were kept running after the activity end, and some sensors were activated in more than one activity, so for getting best feature extraction, we need to remove these records. After removing the noise data, the number of records became 2497 record.

5.2.3 Building Sensors Profile

Sensors profile contain the following fields, *SP = {ID,* $Label_1$, $Label_2$*....*$Label_n$*, Location, Object, Object2, Minimum_interval_time, Maximum_interval_time, Average_interval_time }*

5.2.4 Building Activity Profiles

For Each Activity A has its own profile with the following features: *Ai = {Activity_Label, Location, Sensor_ID, Sensor_Object, Occurrence per activity (%),* *Found in other activities (%), Weight, Max. Sensor IntervalTime, Min. Sensor IntervalTime, Average Activity interval, Primery_Object, Relevant}*

5.2.5 Building Activities List with Location Tag

Activities with Location Tags T contains the following Fields: $T = \{Location\ i, A_1, A_2, \ldots. A_n\}$. the following table shows the location tags for all activities:

Table 10. Location Tags of Activities

Location	**Activity**
Bathroom	Bathing
Bathroom	Cleaning
Bathroom	Doing laundry
Bathroom	Grooming
Bathroom	Putting away groceries
Bathroom	Putting away laundry
Bathroom	Toileting
Bathroom	Washing hands
Bedroom	Bathing
Bedroom	Cleaning
Bedroom	Dressing
Bedroom	Grooming
Bedroom	Putting away laundry
Kitchen	Cleaning
Kitchen	Doing laundry
Kitchen	Going out for shopping
Kitchen	Lawn work

Kitchen	Preparing lunch
Kitchen	Preparing a snack
Kitchen	Preparing a beverage
Kitchen	Preparing breakfast
Kitchen	Preparing dinner
Kitchen	Putting away groceries
Kitchen	Putting away dishes
Kitchen	Washing dishes
Foyer	Going out for shopping
Foyer	Going out to work
Foyer	Going out for entertainment
Living room	Cleaning
Living room	Watching TV
Office/study	Cleaning

5.2.6 Filtering Sensors in Activity Profiles

According to the Location Tag Table (10), and the value of weight in each activity profile, we choose the sensors that their locations refer to that activity or the sensors where their weight value above the midpoint which is (0.5). Other sensors in each profile were removed.

5.2.7 Building Activities Intervals Profile

We divided the day into eight intervals, starting from midnight, each interval has 3 hours. Then for each activity, we found the intervals that the activity-activated in.

Table 11. Activities Intervals

Activity	Interval 1	Interval 2	Interval 3	Interval 4	Interval 5	Interval 6	Interval 7	Interval 8
Bathing	0	0	1	1	1	1	1	0
Cleaning	0	0	0	1	1	0	1	0
Doing laundry	0	0	1	0	1	1	1	0
Dressing	0	0	1	1	1	1	1	1
Going out for entertainment	0	0	0	0	0	0	1	0
Going out for shopping	0	0	0	1	0	0	1	0
Going out to work	0	0	1		1	0	0	0
Grooming	0	0	1	1	1	1	1	1
Lawn work	0	0	0	0	1	0	0	0
Other	0	0	0	0	0	0	1	0
Preparing breakfast	0	1	1	1	0	0	0	0
Preparing dinner	0	0	0	0	0	1	1	0
Preparing lunch	0	0	0	1	1	0	0	0
Putting away dishes	0	0	0	0	1	1	0	0
Putting away groceries	0	0	0	1	0	0	1	0

Putting away laundry	0	0	1	0	1	0	0	0
Toileting	1	1	1	1	1	1	1	1
Washing dishes	0	0	1	1	1	1	1	1
Washing hands	0	0	0	0	0	0	0	1
Watching TV	0	0	1	1	1	1	1	0

5.2.8 Building Sensors Intervals Profile

We did the same as in Activities Intervals, but for sensors. So, for each sensor, we found the intervals that each sensor activated in.

Table 12. Sensors Intervals

Sensor ID	Interval 1	Interval 2	Interval 3	Interval 4	Interval 5	Interval 6	Interval 7	Interval 8
51	0	0	0	1	1	1	0	0
52	0	0	0	1	1	1	1	0
53	0	0	1	1	1	1	1	1
54	0	0	1	1	1	1	1	0
55	0	0	1	1	1	1	1	0
56	0	0	1	1	1	1	1	0
57	1	1	1	1	1	1	1	1
58	1	1	1	1	1	1	1	1
59	0	0	1	1	1	1	0	0
60	0	0	1	1	0	0	0	0
61	0	0	1	1	1	0	1	0

62	0	0	1	1	1	1	1	1
63	0	0	0	1	1	1	1	1
64	0	0	0	0	0	0	1	0
66	0	0	1	1	1	1	1	1
67	0	0	1	1	1	1	1	1
68	1	1	1	1	1	1	1	1
69	0	0	0	0	1	1	0	0
70	0	0	1	1	1	1	1	0
71	0	0	1	0	0	1	1	1
72	0	0	1	1	1	1	1	1
73	0	0	1	1	1	1	1	0
75	0	0	1	1	1	1	1	1
76	0	0	0	0	1	1	1	1
78	0	0	0	1	1	1	1	0
79	0	0	1	1	1	1	1	0
80	0	0	1	1	1	1	1	1
81	0	0	1	0	1	1	1	1
82	0	0	1	1	1	1	1	1
83	0	0	0	1	1	1	1	0
84	0	0	1	1	1	1	1	1
86	0	0	0	1	1	1	1	0
87	0	0	0	1	0	1	0	0
88	1	0	0	0	0	0	0	1
90	0	0	1	1	1	1	1	1

91	0	0	1	1	1	1	1	1
92	0	0	1	1	1	1	1	0
93	0	0	1	1	1	1	1	1
94	0	0	0	1	1	1	1	0
95	0	0	0	1	1	1	1	0
96	1	0	1	1	0	1	1	1
97	0	0	0	1	0	0	0	1
98	0	0	1	0	1	1	1	0
99	0	0	0	0	0	1	0	0
100	0	1	1	1	1	1	1	1
101	1	0	1	1	1	1	1	1
104	0	0	0	0	0	1	1	1
105	0	0	0	1	0	1	1	0
106	0	0	0	0	1	1	1	0
107	0	0	0	0	0	0	1	0
108	0	0	0	0	1	0	0	0
118	0	0	0	0	0	1	0	0
119	0	0	1	1	0	0	0	0
120	0	0	1	1	0	1	1	1
125	0	0	0	0	0	1	0	0
126	0	0	0	1	1	0	1	0
129	0	0	0	1	1	1	1	0
130	1	1	1	1	1	1	1	1
131	0	0	1	1	1	1	1	1

132	0	0	0	1	1	1	0	0
133	0	0	1	0	0	0	0	0
135	0	0	1	1	1	1	1	0
137	0	0	1	1	1	1	1	1
138	0	0	0	1	1	1	0	0
139	0	0	1	1	1	1	1	1
140	0	0	1	1	1	1	1	1
141	0	0	1	1	1	1	1	1
142	0	0	1	1	1	1	1	1
143	0	0	1	1	1	1	1	1
144	0	0	0	1	1	0	0	0
145	0	0	1	1	0	0	0	0
146	0	0	1	1	1	0	0	0
150	0	0	1	1	1	1	1	1

5.3 Data Preprocessing on Target Datasets

Before starting labeling the target domains, we need to do some data preprocessing on them to make it applicable to our approach but without modifying the readings values of the sensors.

5.3.1 Unify Location Names

The layout of the target domains is different from the source domain, so we need to unify the location names to become similar to the source domain. The following table shows the location names of the target domain and how to unify them as the source domain.

Table 13. Sensors Location Names in Target Datasets and its Unified New Name

Location Name	New Name
Bathroom	Bathroom
Bedroom	Bedroom
Kitchen	Kitchen
Living room	Living room
Butler’s Pantry	Kitchen
Dining room	Kitchen
Den	Living room
Hallway	Foyer
Office/study	Office/study

5.3.2 Unify Activities

The target domains contain different activities, but there are some activities not similar to the activities in the source domain, so we removed these activities from the datasets. Also, these activities: Going out to work, Going out for shopping, and Going out for entertainment, are unified to become: Going out, because it is hard to distinguish between them, on another hand, Preparing A snack activity was converted to preparing dinner or lunch or breakfast according to the time of the activity. Table (14), table (15) show the removed activities, and the activities to predict in our approach respectively.

Table 14. Target Domains Activities that not Found in the Source Domain and will not be Labeled

No.	Removed Activities
1	Home education
2	Listening to music
3	Taking medication
4	Talking on telephone
5	Working at computer
6	Sleeping
7	Brush teeth
8	Receive guest
10	Get Drink

Table 15. Target Domains Activities to be Predicted

No.	Activities to Predict
1	Bathing
2	Cleaning
3	Preparing a beverage
4	Grooming
5	Putting away groceries
6	Putting away laundry
7	Toileting
8	Washing hands
9	Dressing
10	Going out
11	Lawn work

12	Preparing lunch
13	Preparing breakfast
14	Preparing dinner
15	Putting away dishes
16	Watching TV

5.4 Fixed Window Size Approach

5.4.1 Segmentation

In fixed size window approach we divided the target domain dataset into segments, each segment has a length less than or equal the average interval time of overall activities in the source domain which is equal = 1123 seconds. Table (16) shows a sample of segments.

Table 16. Sample of segments in the Fixed Window Size Approach

Sensor ID	Sensor_StartTimesec	Sensor_EndTimesec	Segments
137	11341	11386	Segment 1
75	17903	18277	Segment 2
115	17909	17912	Segment 2
74	17923	17930	Segment 2
114	18065	18083	Segment 2
74	18109	18113	Segment 2
114	18125	18127	Segment 2
84	18144	19322	Segment 2
109	18248	20827	Segment 3
….	….	….	….

After applying our code that follows the flow of the segmentation algorithm in Figure (5), we applied our approach as in section (4.3) for detecting the labels of activities. After applying our fixed window size labeling approach, we need to find the percentage accuracy of this approach. For doing this, we need to find the percentage of similarity between the original labels of segments and the detected labels. After segmentation of the target domain, some segments have one activity label, and some have more than one label, for calculating the similarity we need first to make each segment has one label, so for each segment which has more than one label, the label of that segment will be the most frequent label in that segment.

$$\textit{Total Accuracy} = \frac{\textit{No.of correct segments labels}}{\textit{Total No.of segments}} \ldots\ldots\ldots\ldots\ldots \textbf{(3)}$$

5.4.2 Results

After applying Fixed Window Size Approach on the three target datasets , we found the Total Accuracy for all segments and Total Accuracy per activity using equation (3). Table (17) shows the Total Accuracy for the three Target Domain. Tables (18) (19) (20) show the Total Accuracy per activity for MAS S2, OrdonezB, and KasterenA Target Domains respectively.

Table 17. Total Accuracy of Target Domains Using Fixed Window Size Approach

Target Dataset	**MAS S2**	**OrdonezB**	**KasterenA**
No. Of Segments	193	291	186
No. of Corrected Labeled Segments	99	176	147
Total Accuracy	51.29 %	60.48 %	79.03 %

Table 18. Total Accuracy Per Activity Using Window Size Approach on MAS S2 Target Domain

Activity	No. of Labeled Segment	No. of Predicted Segment	Total Accuracy / activity
Bathing	4	3	75 %
Cleaning	4	2	50 %
Dressing	3	1	33.3 %
Going out	4	2	50 %
Grooming	2	2	100 %
Lawn work	2	0	0 %
Preparing breakfast	40	20	50 %
Preparing dinner	40	17	42.5 %
Preparing lunch	25	16	64 %
Putting away dishes	3	1	33.3 %
Putting away groceries	1	0	0 %
Putting away laundry	1	0	0 %
Toileting	27	17	62.3 %
Washing dishes	18	2	11.1 %
Washing TV	19	16	84.2 %

Table 19. Total Accuracy Per Activity Using Window Size Approach on OrdonezB Target Domain

Activity	No. of Labeled Segment	No. of Predicted Segment	Total Accuracy / activity
Bathing	3	0	0 %
Going out	35	15	42.86 %
Grooming	55	40	72.72 %
Preparing breakfast	36	14	38.89 %
Preparing dinner	26	21	80.77 %
Preparing lunch	45	36	80 %
Toileting	24	11	45.83 %
Washing TV	67	39	58.21 %

Table 20. Total Accuracy Per Activity Using Window Size Approach on KasterenA Target Domain

Activity	No. of Labeled Segment	No. of Predicted Segment	Total Accuracy / activity
Bathing	16	16	100 %
Doing laundry	0	0	0 %
Going out	19	7	36.84 %
Preparing breakfast	19	19	100 %
Preparing dinner	36	33	91.67 %

Preparing lunch	4	4	100 %
Putting away dishes	4	0	0 %
Toileting	88	68	77.27 %
Washing dishes	2	0	0 %

5.5 Dynamic Labeling Approach

5.5.1 Approach

In dynamic labeling, segmentation is based on more than one criteria, minimum interval time of overall activities, how much to increase the interval of the segment, and when to move to the next segment.

The minimum activity interval time of overall activity in the source domain is 25 seconds, which is the Toileting activity. So firstly, each segment will have an interval length less than or equal 25 seconds, after detecting the similarity between sensors and detecting the label for that segment as in fixed size window labeling, we increase the length of that segment by value of the minimum interval time for that activity from the activities profiles, so this will be the lower limit of the segment. We move to the next segment if the sensor different between sensor start time and lower limit of the segment is larger than or equal the double value of minimum interval time.

After applying our code to all records in the target datasets, each record has its corresponding detected label.

To find the percent of similarity between the detected labels and the original activities label for the target domain, we applied the following equation:

$$\boldsymbol{Total\ Accuracy} = \frac{\boldsymbol{No.of\ correct\ Records\ labels}}{\boldsymbol{Total\ No.of\ records}} \ \textbf{.........................} \ \textbf{(4)}$$

5.5.2 Results

After applying Dynamic Labeling Approach on the three target datasets, we found the Total Accuracy for all Datasets Records and Total Accuracy per activity using equation (4). Table (21) shows the Total Accuracy for the three Target Domain. Tables (22) (23) (24) show the Total Accuracy per activity for MAS S2, OrdonezB, and KasterenA Target Domains respectively.

Table 21. Total Accuracy of Target Domain Using Dynamic Labeling Approach

Target Dataset	**MAS S2**	**OrdonezB**	**KasterenA**
No. Of Records	1513	1082	873
No. of Corrected Labeled Records	1095	816	697
Total Accuracy	72.37 %	75.42 %	77.78 %

Table 22. Total Accuracy Per Activity Using Dynamic Labeling Approach on MAS S2 Target Domain

Activity	**No. of Labeled Records**	**No. of Predicted Records**	**Total Accuracy / Activity**
Bathing	224	214	95.54 %
Cleaning	25	4	16.00 %
Dressing	16	4	25.00 %
Going out	9	7	77.78 %
Grooming	7	2	28.57 %
Lawn work	2	0	0 %

Preparing breakfast	376	274	72.87 %
Preparing dinner	281	225	80.07 %
Preparing lunch	217	170	78.34 %
Putting away dishes	35	0	0 %
Putting away groceries	13	0	0 %
Putting away laundry	3	0	0 %
Toileting	145	118	81.38 %
Washing dishes	79	0	0 %
Washing TV	81	77	95.06 %

Table 23. Total Accuracy Per Activity Using Dynamic Labeling Approach on OrdonezB Target Domain

Activity	No. of Labeled Records	No. of Predicted Records	Total Accuracy / Activity
Bathing	10	10	100 %
Going out	82	50	60.98 %
Grooming	160	110	68.75 %
Preparing breakfast	165	160	96.97 %
Preparing dinner	127	125	98.43 %

Preparing lunch	197	192	97.46 %
Toileting	120	78	65.00 %
Washing TV	221	91	41.18 %

Table 24. Total Accuracy Per Activity Using Dynamic Labeling Approach on KasterenA Target Domain

Activity	No. of Labeled Records	No. of Predicted Records	Total Accuracy / Activity
Bathing	52	50	96.15 %
Doing laundry	13	0	0 %
Going out	61	43	70.49 %
Preparing breakfast	127	106	83.46 %
Preparing dinner	195	174	89.23 %
Preparing lunch	19	3	15.79%
Putting away dishes	26	0	0 %
Toileting	368	314	85.32 %
Washing dishes	12	7	58.33 %

5.6 Discussion

According to Fixed Size Window Approach, the reason for the low accuracy in MAS S1 and OrdonezB target datasets is due to the occurrence of some different activities on the same day and within a short period of time. As the period of time between an activity and the next activity is less than the average interval time of activities as a whole and this led to the overlap of some activities with each other in some segments, which led to the prediction of an activity label similar to one of the activities located in the segment or the prediction of an activity label not similar to existing activities. So, Where the presence of more than one activity in the same segment, we have taken the most repetitive activity in the segment and this leads to an error in determining the original activity in a single segment, which leads to a reduction in the total accuracy ratio. Another reason for overlapping activities within the same segment, some activities have an interval time larger than the average interval time of overall activities, so this will cause that the activity which has large interval time will be divided into one or more than one segment, so part of this activity may move to another segment with different label.

Some activities in both approaches have low total accuracy or even 0 % total accuracy, the reason for the low value of their accuracy is due to one of the following reasons:

1- Some activities contain a good number of sensor readings that are not relevant to them and which can be closer to other activities. this causes a label to be predicted for these activities will not be similar to the original activity label.

2- Some activities also contain sensor readings that are not completely related to these activities. This is due to errors in the original data, possibly due to faults in the sensors themselves. This also leads to the prediction of wrong labels for these activities.

3- Some activities are similar in nature of doing them to other activities and also contain similar sensor readings. For example, doing dishwashing or washing clothes results to activate the sensors installed on these machines or tools that are used for doing these activities, also when putting away the dishes or clothes, the same sensors will be activated, so it will be difficult to distinguish between these activities and determine the appropriate label for them.

4- A short period of time between activities may also lead to incorrect labeling, so during labeling the activities, one activity may have the same label name with previous activity. Also, in Dynamic Labeling Approach, we use Minimum Activity Interval (which is a feature extracted from the source domain) to expand the segment size during the labeling process, so if an activity has been ended within a short time less than the minimum activity interval time of the predicted activity, the next activity or part of its records will be labeled with same label name of the previous one.

6. Conclusion and Future Works

Activity recognition is a central issue in smart homes, different in homes layout and sensors distribution in smart homes, and even different between individual behaviors, request huge amount of data collecting and analyzing in order to be able to carry out the conventional activity discovery and recognition algorithms, and then, it will need lots of time in order to learn the new environments. Transfer learning techniques come to be a good solution to handle these problems, we transfer the knowledge acquired from the source domain to the target domain in order decrease the time spent in adapting new tasks and deal with more circumstances adequately and making the educated model more robust.

We applied two approaches in this work, Fixed Window Size Approach, where the segmentation and labeling the data based on fixed time interval which is the average interval time for overall activities, and Dynamic Labeling Approach, where the segmentation and labeling depend on the interval time of the detected label, so the segments length will be different depend on the detected label. Fixed Window Size

Approach shows good results when the activities have time period between them larger than the window size of segmentation. Dynamic Approach has a best results of overall accuracy for the three target domains and can deal with various circumstances. But the two approaches could not deal well with activities that have the same nature of work.

In our future work, we will enhance our approaches to deal with more circumstances and to deal with activities that have same nature of work to increase the accuracy of the results, on other hand, we will increase the number of activities to recognize and predict more human daily activities to create a general transfer model for smart homes.

References

Alliance, F. C., 2017. *Family Caregiver Alliance.* [Online] Available at: https://www.caregiver.org/selected-long-term-care-statistics [Accessed 22 10 2017].

Blanke, U. & Schiele, B., 2010. *Remember and transfer what you have learned-recognizing composite activities based on activity spotting.* In *Wearable Computers (ISWC), 2010 International Symposium on* (pp. 1-8). IEEE.

Blitzer, J., McDonald, R. & Pereira, F., 2006. *Domain adaptation with structural correspondence learning.* In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 120-128). Association for Computational Linguistics.

Byrnes, J. P., 1996. *Cognitive development and learning in instructional contexts.* Boston: Allyn and Bacon.

Cao, L., Liu, Z. & Huang, T. S., 2010. *Cross-dataset action detection.* pp. 1998-2005.

Chen, W.-H., Cho, P.-C. & Jiang, Y.-L., 2017. Activity Recognition Using Transfer Learning. *Sensors and Materials,* Volume 29, pp. 897-904.

Chiang, Y.-t. & Hsu, J. Y.-j., 2012. *Knowledge Transfer in Activity Recognition Using Sensor Profile.* In *Ubiquitous Intelligence & Computing and 9th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2012 9th International Conference on* (pp. 180-187). IEEE.

Cook, D. J., 2012. Learning setting-generalized activity models for smart spaces. *IEEE intelligent systems,* Volume 27, pp. 32-38.

Daume III, H., 2009. Frustratingly easy domain adaptation. *arXiv preprint arXiv:0907.1815.*

Diethe, T., Twomey, N. and Flach, P., 2015, July. Bayesian active transfer learning in smart homes. In *ICML Active Learning Workshop* (Vol. 2015)..

Diethe, T., Twomey, N. and Flach, P., 2016, April. Active transfer learning for activity recognition. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*..

Dillon Feuz, K. & J. Cook, D., 2014. Heterogeneous transfer learning for activity recognition using heuristic search techniques. *International Journal of Pervasive Computing and Communications,* Volume 10, pp. 393-418.

Fahad, L., A., K. & M., R., 2015. Activity Recognition in smart homes with self verification of assignments. *Neurocomputing,* Volume 149, pp. 1286 - 1298.

Fang, H. & Hu, C., 2014. Recognizing human activity in smart home using deep learning algorithm. *Control Conference(CCC),* pp. 4716-4720.

Feuz, K. D. & Cook, D. J., 2015. Transfer learning across feature-rich heterogeneous feature spaces via feature-space remapping (FSR). *ACM Transactions on Intelligent Systems and Technology (TIST),* Volume 6, p. 3.

Fleury, A., Vacher, M. & Noury, N., 2010. SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results. *IEEE transactions on information technology in biomedicine,* Volume 14, pp. 274-283.

Freund, Y. & Schapire, R., 1995. A desicion-theoretic generalization of on-line learning and an application to boosting. *European conference on computational learning theory,* pp. pp.23-37.

Hachiya, H., Sugiyama, M. & Ueda, 2012. Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition. *Neurocomputing,* Volume 80, pp. pp.93-101.

Hossain, H. M. S., Khan, M. A. A. H. & Roy, N., 2017. Active learning enabled activity recognition. *Pervasive and Mobile Computing,* Volume 38, pp. 312-330.

Hu, C. et al., 2018. A Novel Feature Incremental Learning Method for Sensor-based Activity Recognition. *IEEE Transactions on Knowledge and Data Engineering.*

Hu, D. and Yang, Q., 2011. Transfer learning for activity recognition via sensor mapping. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain* (p. 1962).

Inoue, S. and Pan, X., 2016, November. Supervised and Unsupervised Transfer Learning for Activity Recognition from Simple In-home Sensors. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (pp. 20-27). ACM.

Kasteren, T. & Krose, B., 2007. Bayesian activity recognition in residence for elders.

Kasteren, T. L. M., Englebienne, G. & Krose, B. J. A., 2011. Human activity recognition from wireless sensor network data: Benchmark and software. In: *Activity recognition in pervasive intelligent environments.* s.l.:Springer, pp. 165-186.

Liu, F., Zhang, G., Lu, H. & Lu, J., 2017. Heterogeneous Unsupervised Cross-domain Transfer Learning. *CoRR.*

Mannini, A. & Sabatini, A. M., 2010. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors,* Volume 10, pp. 1154-1175.

Munguia Tapia, E., 2003. *Activity recognition in the home setting using simple and ubiquitous sensors* (Doctoral dissertation, Massachusetts Institute of Technology).

Ordonez, F. J., Toledo, P. & Sanchis, A., 2013. Activity recognition using hybrid generative/discriminative models on home environments using binary sensors. *Sensors,* Volume 13, pp. 5460-5477.

Pan, S. J. & Yang, Q., 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering,* Volume 22, pp. 1345-1359.

Parkka, J., Cluitmans, L. & Ermes, M., 2010. Personalization algorithm for real-time activity recognition using PDA, wireless motion bands, and binary decision tree. *IEEE Transactions on Information Technology in Biomedicine,* Volume 14, pp. 1211-1215.

Pratt, S. T. ., 1998. *Learning to learn. Kluwer, Berlin.*

Rashidi, P. and Cook, D.J., 2010, January. Activity Recognition Based on Home to Home Transfer Learning. In *Plan, Activity, and Intent Recognition*.

Rashidi, P. & Cook, D. J., 2010. Multi home transfer learning for resident activity discovery and recognition. *KDD Knowledge Discovery from Sensor Data*, pp. 56-63.

Rawashdeh, M. et al., 2017. A knowledge-driven approach for activity recognition in smart homes based on activity profiling. *Future Generation Computer Systems.*

Samarah, S. et al., 2018. Transferring activity recognition models in FOG computing architecture. *Journal of Parallel and Distributed Computing,* Volume 122, pp. 122-130.

Van Kasteren, T.L.M., Englebienne, G. and Kröse, B.J., 2010, May. Transferring knowledge of activity recognition across sensor networks. In *International Conference on Pervasive Computing* (pp. 283-300). Springer, Berlin, Heidelberg.

Venkatesan, A., 2011. *A study of boosting based transfer learning for activity and gesture recognition.* s.l.:Arizona State University.

Ying, J.J.C., Lin, B.H., Tseng, V.S. and Hsieh, S.Y., 2015, October. Transfer learning on high variety domains for activity recognition. In *Proceedings of the ASE BigData & SocialInformatics 2015* (p. 37). ACM.

الخطيب، خالد . تقنية نقل التعلم للتعرف على الانشطة في المنازل الذكية. ماجستير في نظم المعلومات الحاسوبية، رسالة، قسم نظم المعلومات الحاسوبية، جامعة اليرموك، 2018. (المشرف: أ.د. سامر سمارة)

الملخص

تقنية نقل التعلم للتعرف على الانشطة في المنازل الذكية

يعتبر التعرف على النشاط البشري مهمة أساسية في البيوت الذكية، حيث يعد اكتشاف الأنشطة البشرية مجموعة من التقنيات التي يمكن استخدامها في نطاق واسع من التطبيقات، ويعتمد على تقنيات الذكاء الاصطناعي لفهم بيانات المستشعر وللحصول على استخدام المعلومات لأنشطة التعرف والتتبع. ومع ذلك، تم تصميم العديد من التقنيات التي تم تطويرها لحالات مبسطة. يتمثل أحد التحديات في التعرف على النشاط في الحاجة إلى جمع كميات كبيرة من البيانات وتوضيحها حتى يكون بمقدورنا تنفيذ خوارزميات اكتشاف وتمييز الأنشطة التقليدية. هذه المرحلة الأولية من جمع البيانات وتحليلها ينتج عنها عملية تثبيت مطولة واستثمار مدة زمنية طويلة لكل مساحة ومنزل جديدين. يتناول نقل التعلم مشكلة كيفية الاستفادة من المعرفة المكتسبة مسبقًا (المصدر) لتحسين كفاءة التعلم في مجال جديد (المجال المستهدف) واستخدامه كحل لمعالجة قضايا التنوع للتعرف على النشاط في المنازل الذكية. في هذه الرسالة، نقترح طريقة جديدة لنقل المعرفة المكتسبة من الأنشطة إلى مساحة جديدة من أجل الاستفادة من عملية التعلم في البيئة الجديدة. لقد استخدمنا نهجين في العمل، أولا) نهج التسمية من خلال إطار نافذة ثابت للبيانات، حيث يتم تقسيم البيانات ووضع علامات عليها استنادًا إلى فترة زمنية محددة وهي متوسط الفترة الزمنية للأنشطة العامة ، ثانيا) نهج تسمية ديناميكية حيث تعتمد التجزئة و التقسيم على زمن النشاط المستكشف ، وبالتالي فإن طول المقاطع ستكون مختلفة تعتمد على الانشطه المكتشفة. للتحقق من صحة خوارزمياتنا، استخدمنا مجوعة من البيانات التي تم جمعها في العديد من البيئات الذكية ذات التخطيطات والتقسيمات المختلفة وأيضا ذات سلوكيات الأفراد المختلفة. يعرض أسلوب حجم الإطار الثابت نتائج جيدة عندما يكون للأنشطة وقتًا زمنيًا بينها أكبر من حجم نافذة التجزئة.بينما يتمتع النهج الديناميكي

بأفضل نتيجة للدقة الشاملة للنطاقات الثلاثة المستهدفة ويمكنه التعامل مع مختلف الظروف. لكن لا يمكن للطريقتين التعامل بشكل جيد مع الأنشطة التي لها نفس طبيعة العمل.