Automated Cognitive Health Assessment Using Smart Home Monitoring of Complex Tasks

Prafulla N. Dawadi, *Student Member*, *IEEE,* Diane J. Cook, *IEEE Fellow*, and Maureen Schmitter-Edgecombe

***Abstract*— One of the many services that intelligent systems can provide is the automated assessment of resident well-being. We hypothesize that the functional health of individuals, or ability of individuals to perform activities independently without assistance, can be estimated by tracking their activities using smart home technologies. In this paper, we introduce a machine learning-based method for assessing activity quality in smart homes. To validate our approach we quantify activity quality for 179 volunteer participants who performed a complex, interweaved set of activities in our smart home apartment. We compare our automated assessment of task quality with direct observation scores. We also assess the ability of machine learning techniques to predict the cognitive health of the participants based on these automated scores. We believe that this capability is an important step in understanding everyday functional health of individuals in their home environments.**

***Index Terms*— Smart environments, Machine learning, Automated assessment**

I. Introduction

The maturing of ubiquitous computing technologies has opened the doors for application of these technologies to areas of critical need. One such area is ubiquitous monitoring of an individual's cognitive and physical health. The possibilities of using smart environments for health monitoring and assistance are perceived as "extraordinary" [11] and are timely given the aging of the population [4][5][41][43].

We hypothesize cognitive impairment can be evident in everyday task performance. We also postulate that differences in task performance can be automatically detected between cognitively healthy individuals and those with dementia and mild cognitive impairment (MCI) using smart home and ubiquitous computing technologies. In one of the first projects to focus on this question, we investigate approaches for quantifying task performance and relate the automated scores to cognitive health of individuals.

This material is supported by the Life Sciences Discovery Fund, by National Science Foundation grants 1064628 and 0852172, and by the National Institutes of Health grant R01EB009675.

Prafulla Dawadi and Diane Cook are with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, 99164 (e-mail: prafulla.dawadi@wsu.edu , djcook@wsu.edu).
Maureen Schmitter-Edgecombe is with the Department of Psychology, Washington State University, Pullman, WA, 99164 (e-mail: schmitter-e@wsu.edu).

Neuropsychologists and clinicians are interested in understanding everyday functioning of individuals to gain insights about difficulties that affect quality of life and to assist individuals in completing daily activities. Everyday functioning encompasses a range of daily functional abilities such as cooking, managing finances, driving, and other activities of daily living that individuals must complete to live competently and independently. In addition, deficits and changes in everyday functioning are considered as precursors to more serious cognitive problems such as dementia and MCI [13][44]. As proxy measures for everyday functioning, clinicians typically use performance-based simulation measures administered in a laboratory setting and/or self-report and informant-report questionnaires of activities of daily living. Though these methods are thought to reflect activity performance in realistic settings, the assessment techniques are widely questioned for their ecological validity [7]. For example, self-report and informant-report are subject to reporter bias while data that is collected via simulation measures in a lab or clinical setting may not capture subtle details of activity performance that occur in a home setting [31]. Among these methods, direct observation of the individual to determine everyday functional status is widely considered the most ecologically valid [7] [34].

When observing individuals performing everyday activities, clinicians can derive information about how well the individual is able to perform the activities. If important steps are skipped or performed incorrectly then the benefit of the activity is not realized and the mistake may be indicative of a health condition. Such activity mistakes may include forgetting to turn off the burner, keeping the refrigerator door open, or taking an unusually long time to complete a relatively simple activity. When activities can be assessed in an individual's own environment, they can help in characterizing daily behavior and pinpointing cognitive or physical difficulties.

In this paper, we propose a machine learning methodology to automatically quantify the quality of the performance of an activity with respect to how other individuals perform the same activity. We implement our approach to activity assessment in our smart home testbed and correlate our automated scores with measurements derived from direct observation of participant performances. Finally, we analyze correlation between activity quality and health diagnosis for these individuals.

II. Related Work

A smart home can be viewed as an environment in which computing and communications technologies employ artificial intelligence techniques to reason about and control our physical home setting [9]. In a smart home, sensor events are generated while residents perform their normal daily routines. A smart home is an ideal environment for performing automated health monitoring and assessment. Using this setting no constraints are made on the resident's lifestyle. As an example, Pavel et al. [26] hypothesized that change in mobility patterns are related to change in cognitive ability. They tested this theory by observing changes in mobility as monitored by motion sensors and found evidence to support the relationship between these changes and symptoms of cognitive decline. Lee and Dey [19] also designed an embedded sensing system and presented information to older adults to determine if this information was useful to them in gaining increased awareness of their functional abilities.

The ability to perform automated assessment of task quality and cognitive health has recently been given a boost because activity recognition techniques are becoming more capable of accurately identifying the current task [8][17]. These techniques map a sequence of readings from a particular sensor modality to a label indicating the activity that is being performed. In our experiments, we rely upon environmental sensors including infrared motion detectors and magnetic door sensors that are useful for gathering information about complex activities such as cooking, sleeping, and eating [3][18][42]. However, the techniques we describe in this paper can just as easily make use of other sensor data. Possibilities include wearable sensors such as accelerometers, which are commonly used for recognizing ambulatory movements (e.g., walking, running, sitting, climbing, and falling) [22]. More recently, researchers are exploring smart phones equipped with accelerometers and gyroscopes to recognize such movement and gesture patterns [18]. Some activities such as washing dishes, taking medicine, and using the phone are characterized by interacting with unique objects. In response, researchers have explored the usage of RFID tags and shimmer sensors for tagging these objects and using the data for activity recognition [23][29]. Other researchers have used data from video cameras and microphones as well [3]. In another automated assessment effort, Allin and Ecker [2] used computer vision techniques to correlate motor statistic of stroke survivor's motion obtained from multiple digital cameras with expert functional scores on the Arm Motor Ability Test (AMAT).

While smart environment technologies have been studied extensively for the purposes of activity recognition and context-aware automation, less attention has been directed toward using the technologies to assess the quality of tasks performed in the environment. Some earlier work has measured activity correctness for simple [15] and strictly sequential tasks. Cook and Schmitter-Edgecombe [10] developed a model to assess the completeness of activities using Markov models. Their model detected certain types of step errors, time lags, and missteps. Similarly, Hodges et al. [15] correlate sensor events gathered during a coffee-making task with an individual's neuropsychological score. They found a positive correlation between sensor features and the first principal component of the standard neuropsychological scores. In another effort, Rabbi et al. [31] designed a sensing system to assess mental and physical health using motion and audio data.

In contrast to these other projects, we are analyzing data from parallel and interwoven activities and are correlating sensor features with scores derived from direct observation of performance on a complex task. This represents one of the first reported projects to use smart home technologies to automate such assessment for a large group of participants.

III. Background

The primary goal of this project is to perform automated assessment of tasks performed in a smart home. Given a sequence of sensor events that are generated while an activity is being performed, our algorithm will map the sequence onto a quantitative score. We assume that sufficient sensors exist to capture activity progression in detail and that the smart home captures the timing for each sensor event. In this project, we also assume that activities are complex with a number of independent sub-activities that may be interwoven with steps from other activities or sub-activities. This situation represents a challenge that recently has been addressed by the activity recognition community [30][37][45].

A. The Testbed

Data is collected and analyzed using the Washington State University CASAS on-campus smart home testbed, an apartment that contains a living room, a dining area, and a kitchen on the first floor and two bedrooms, an office, and a bathroom on the second floor. The apartment is instrumented with motion sensors on the ceiling, door sensors on cabinets and doors, and item sensors on selected kitchen items. The testbed contains temperature sensors in each room, sensors to monitor water and burner use, and a power meter to measure electricity consumption. Item sensors are placed on a set of items in the apartment to monitor their use. Figure 1 shows the sensor layout in the CASAS smart home testbed. All of the activities for this study were performed in the downstairs of the apartment while an experimenter monitored the participant upstairs via a web camera and remotely communicated to the participant using a microphone and a speaker.

Sensor events are generated and stored while participants perform activities. Each sensor event is represented by four fields: *date, time, sensor identifier, and sensor message.* The data files and corresponding video are examined by humans who annotate the data with the activity that is being performed that causes the sensor event to be generated. A sample of the collected sensor events, together with the corresponding labeled activities, is shown in Figure 2. CASAS middleware collects sensor events and stores the data in an SQL database. All software runs locally using a small Dream Plug computer.

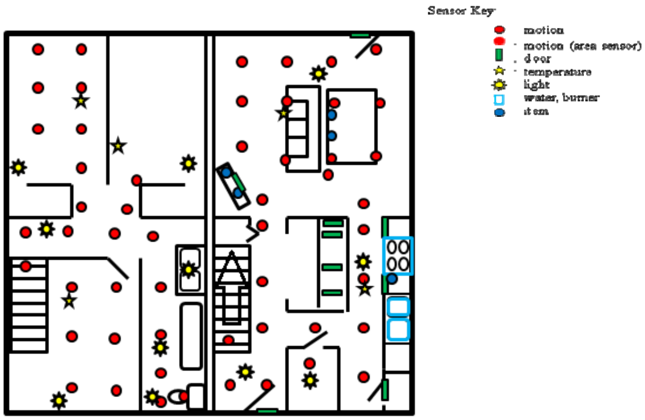

Fig. 1. Apartment testbed floor plan and sensor layout.

Formally, the input data to our algorithm is a sequence of sensor events, *E*, that is generated as an individual performs an activity, A, which is comprised of subtasks $A_1..A_n$. A subtask A_i is represented by the corresponding sequence of n sensor events $e_1..e_n$, the start time of the activity, the end time of the activity, and the activity label. Activity subtasks can be initiated in an arbitrary order and some activities or activity subtasks can be interwoven or parallelized. We state that activity A at sensor event e_i is parallelized if there is more than one subtask open (started but not ended) at that time.

Date Time	ID		Message
2010-01-29 15:32:44.37929	M013	ON	Bus/Map - end
2010-01-29 15:32:44.40339	M010	ON	Change - start
2010-01-29 15:32:48.35166	M010	OFF	
2010-01-29 15:32:50.25128	T006	23	
2010-01-29 15:32:51.47531	M014	ON	
2010-01-29 15:32:51.49961	M013	ON	
2010-01-29 15:33:00.2511	M013	OFF	
2010-01-29 15:33:13.08106	P001	1108	
2010-01-29 15:33:31.84137	M015	ON	Change - end
2010-01-29 15:33:36.11199	M016	OFF	
2010-01-29 15:33:58.00101	M017	ON	Magazine - end

Fig. 2. Sensor file format and sample annotation. Sensor IDs starting with M are motion sensors, D are door sensors, T are temperature sensors, and P are power usage sensors. The data is annotated with the start and end points of the subtasks. The sample annotation shows a participant interweaving the Magazine, Bus/Map, and Change tasks.

B. *The Day Out Task*

The ability to multi-task, or perform concurrent tasks or jobs by interleaving, has been said to be at the core of competency in everyday life [6]. We therefore designed a "Day Out Task" (DOT), a naturalistic task that participants complete by interweaving subtasks. Participants were told to imagine that they were planning for a day out, which would include meeting a friend at a museum at 10am and later traveling to the friend's house for dinner. The eight subtasks that need to be completed to prepare for the day out are explained and participants are told to multi-task and perform steps in any order to complete the preparation as efficiently as possible. Participants are also provided with a list and brief description of each subtask that they can refer to during DOT completion.

The eight subtasks are:

- *Magazine*: Choose a magazine from the coffee table to read on the bus ride.
- *Heating pad*: Microwave for 3 minutes a heating pad located in the kitchen cupboard to take on the bus.
- *Medication*: Right before leaving, take motion sickness medicine found in the kitchen cabinet.
- *Bus map*: Plan a bus route using a provided map, determine the time that will be needed for the trip and calculate when to leave the house to make the bus.
- *Change*: Gather correct change for the bus.
- *Recipe*: Find a recipe for spaghetti sauce in the recipe book and collect ingredients to make the sauce with your friend.
- *Picnic basket*: Pack all of the items in a picnic basket located in the closet.
- *Exit*: When all the preparations are made, take the picnic basket to the front door.

C. *Experimental Setup*

Data from 179 participants was used to validate our approach for activity assessment. Participants initially completed standardized and experimental neuropsychological tests in a laboratory setting before attempting the DOT task in our smart home testbed. In this paper, we analyze data from cognitively healthy participants (N=145) as well as those with dementia (N=2) and mild cognitive impairment (N=32) who completed at least two of the eight subtasks. Participants in the dementia group met DSM-IV-TR criteria for dementia [5], which includes the presence of multiple cognitive deficits that negatively affect everyday functioning and represent a decline from a prior level of functioning. Inclusion criteria for MCI were consistent with the diagnostic criteria defined by Petersen [27][28] and with the criteria outlined by the National Institute on Aging-Alzheimer's Association workgroup [1]. The participant pool included 141 females and 38 males, with 37 participants under 45 years of age (YoungYoung), 27 participants age 45-59 (MiddleAge), 84 participants age 60-74 (YoungOld), and 31 participants age 75+ (OldOld).

While participants were completing the DOT, two examiners remained upstairs in the apartment, watching participant performances through live feed video. As participants completed the DOT, the examiners recorded the time each subtask began and ended, events being interweaved, and subtasks goals being completed (e.g., retrieves magazine). As the individuals perform activities in the smart home, sensors generate events that are recorded. A research team member annotated the sensor data to relate events with the label of the subtask that the individual was performing when the event was triggered. Figure 2 shows a sample of the collected and annotated sensor data. Subtask accuracy scores and task sequencing scores were later assigned by coders after watching the video. Figure 3 illustrates this process.

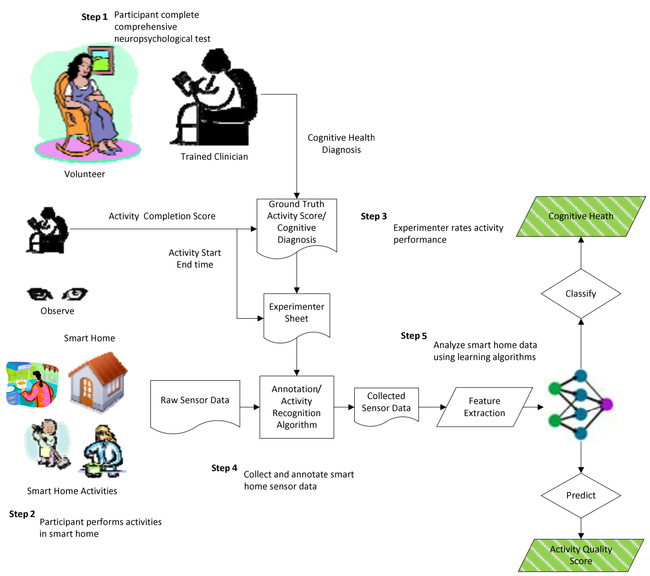

Fig. 3. Automated task assessment steps.

D. *Task Scoring*

Our objective is to automatically quantify task quality as it is performed by a participant. To provide a basis of comparison for the sensor data, two trained neuropsychologists watched the video data and, in conjunction with examiner-recorded data, assigned a task accuracy score and a sequencing score. The task *accuracy score* was based on the correctness and completeness of each of the eight subtasks that the participant needed to complete. A correct and complete subtask received a lower score while an incorrect, incomplete, or uninitiated subtask received a higher score. The scoring criteria are listed in Tables 1 and 2. The final accuracy score was obtained by summing the individual scores of each task and thus ranged from 8 to 32. The task *sequencing score* represents whether the participant sequenced six of the DOT subtasks correctly. Participants received 1 point for each correct sequence (e.g., put the heating pad in the microwave for 3 minutes as one of the first four subtasks; looked up the recipe prior to beginning to retrieve food items). The normalized range of scores is 1 to 6 such that lower score indicates a more correct and/or efficient sequencing of subtasks. Inter-rater reliability was high, with greater than 95% agreement between raters for both the task accuracy and sequencing scores. Figure 4 shows the distribution of the direct observation scores, accuracy and sequencing score grouped by participant's cognitive diagnosis. Refer to [35] for more details on task scoring and DOT.

IV. Feature Extraction

To assess an individual's performance on the DOT, we derive features from annotated sensor data that reflect task performance and can be fed as input to a machine learning algorithm to quantify task quality. We define DOT performance based on two concepts: nature of activity completion and execution of the activity subtasks. A participant efficiently executes DOT if he multitasks DOT subtasks and sequences them correctly. Similarly, time taken to complete the entire DOT activity and number of sensors triggered during activity completion explains the participant's nature of DOT completion. Based on these concepts of activity performance, we derive a set of DOT features. DOT features analyze DOT activity from several perspectives such as how long the participant took to complete the entire DOT and the amount of subtask parallelism / interweaving they employed. We note that in this study the activity start points and end points were generated by human annotators. However, with the use of activity recognition algorithms this step can be automated as well [37]. The feature vector describing one individual's performance of the DOT consists of the activity duration, sensor event counts, degree of task parallelism, number of activities completed, pattern sequencing, and number of activity interruptions. These features were chosen because prior studies have found that, in comparison to cognitively healthy older adults, individuals with MCI complete everyday activities (e.g., locating nutrition information on food labels, making toast, medication management) more slowly [44] and commit more errors, including errors of commission [14], omission [24], and task sequencing/tracking [36].

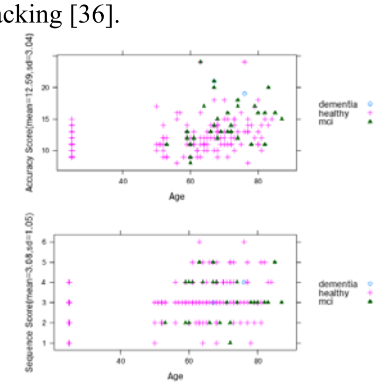

Fig. 4. Distribution of the neuropsychologist direct observation scores, accuracy score and sequencing score, grouped by participant's cognitive diagnosis. Individual participants are organized by age on the x axis and by the corresponding score on the y axis.

TABLE 1
CODING SCHEME TO ASSIGN ACCURACY SCORE TO EACH SUBTASK

Accuracy score	Criteria
1	Complete / Efficient
2	Complete / Inefficient
3	Incomplete / Inaccurate
4	Never Attempted

TABLE 2
CODING SCHEME TO ASSIGN SEQUENCING SCORE TO EACH SUBTASK. TOTAL SEQUENCING SCORE IS THE COUNT OF "YES" RESPONSES TO THESE CRITERIA.

ID	Criteria
1	Heating pad started as one of first four activities.
2	Picnic basket retrieved as one of first four activities.
3	Cost of bus fare determined prior to first attempt at retrieving change.
4	Recipe read prior to retrieving first food item.
5	Dramamine pill taken near end.
6	Picnic basket moved to front door as one of last two activities.

A. Duration

We use the duration feature to represent the total wall clock time that the participant takes to complete the entire set of DOT activities. The time that an individual takes to complete an activity can be indicative of the participant's age as well as mobility and overall cognitive health. If subtasks are executed independently then we can consider the time for each subtask as a separate feature. For the DOT subtasks are interleaved and performed in parallel, so we only consider time taken for the entire DOT.

B. Number of Sensors and Sensor Events

This feature reflects the spatial areas and objects that are manipulated while DOT is being performed. The number of sensors indicates the number of different sensor identifiers that generate events during the DOT, while the number of sensor events keeps track of the number of events that is generated by each unique sensor in the space. These counts provide insight on the type of activities that are being performed and how well the participant stays on the task. For example, some participants wandered out of the normal activity region, used incorrect tools for a subtask, or explored the same space, cabinet or region repetitively as they attempted to complete the appropriate subtask.

B. Parallelism

Participants in our study were encouraged to multitask the DOT subtasks as much as possible to complete the DOT quickly. The ability to multitask varied dramatically among individuals and was expected to present a challenge for those with dementia and MCI. We were therefore interested in quantifying the amount of parallelism or multitasking that existed in an individual's performance of the DOT.

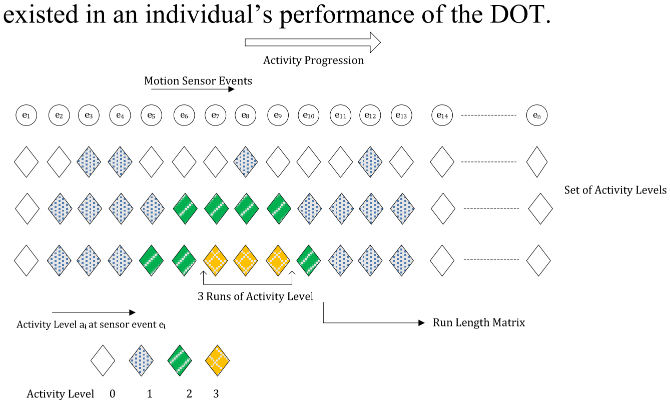

Fig. 5. Sets of activity levels for three participants. The first item of the set represents activity level at the initial sensor event. As activity progresses, sets are augmented with activity levels for different sensor events. For example, during the eighth sensor event *e8*, participant III has activity level 3 and participant I has 1. The run length matrix takes this activity level set as input.

To quantify parallelism, we introduce a variable called activity level, a_i, that represents the number of activities that are open (i.e., that have been started but not completed), at the time that sensor event i is generated. A set of activity levels $\{a_1, a_2, .., a_n\}$ can be defined for all of the sensor events, $1..n$, that were generated while an individual completed the DOT. To represent this set more succinctly we employ run length encoding (RLE). A run for an activity level is a string of equal-valued activity levels. RLE encodes runs of activity levels as activity levels with corresponding counts. The process is explained in Figure 5. Based on the run length encoding, we derive a run length matrix P of size $M \times N$, where M is the maximum activity level and N is the length of the sensor sequence. Each element of the matrix, $P(x,y)$, represents the number of runs of length y corresponding to activity level x, or the number of times that activity level x occurs y consecutive times. A similar technique has been used in the context of medical image analysis to analyze computed tomography volumetric data to capture various text characteristics [12].

We introduce two measures, the High Activity-Level Run Measure (HALRM) and the Low Activity-Level Run Measure (LALRM), to capture a participant's level of task parallelism that occurred over a sequence of sensor events. If a participant parallelizes subtasks for a longer period of time we expect his HALRM to be high, while if he does not parallelize subtasks his LALRM measure would be high.

$$HALRM = \sum_{i=1}^{M} \sum_{j=1}^{N} P(i,j) \times i \times j \tag{1}$$

$$LALRM = \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P(i,j) \times j}{i} \tag{2}$$

Based on these two measures, we define the parallelizing index, Pindex, to represent the amount of task interweaving that is performed. Pindex is computed as the ratio of HALRM to LALRM, as shown in Equation 3.

$$Pindex = \frac{HALRM}{LALRM} \tag{3}$$

As Equation 3 indicates, a higher parallelizing index indicates a higher level of parallelism in the activity. It does not reflect a higher quality of DOT. For example, a participant may have a high *Pindex* because he initiated many of the subtasks. On the other hand, he may leave subtasks incomplete or take a long time to complete the subtasks. The *Pindex* does provide particularly useful insights on task quality when combined with the other task features.

C. Number of Complete Activities

Not all participants complete all DOT subtasks. We therefore introduce a Boolean feature, *ActivitiesCompleted*, which indicates whether the participant completed all of the activity subtasks (in this case, the DOT subtasks).

D. Pattern Sequencing

In the case of a complex activity such as the DOT, subtasks can be performed with many order variations. For instance, one participant might choose a magazine first, while another might start by first looking up a recipe. Participants are expected to parallelize subtasks for efficiency. However, some subtask sequences and parallelisms are more efficient than others. As an example, if a participant starts the DOT by microwaving a heating pad, they are able to complete other tasks while waiting for the microwave to finish. If they wait until the end of the DOT to microwave the heating pad this parallelism is not possible. We hypothesize that the sequence in which tasks are performed influences the amount of parallelism that can be achieved and thereby affects the efficiency of the overall task.

To represent task sequencing choices, we define a DOT sequencing vector $s_1, s_2, .., s_8$ that encodes the order in which an individual started various tasks (in the DOT, there are 8 such tasks to choose from). For example, the sequencing vector (2, 3, 1, 4, 5, 6, 7, 8) indicates that the 2nd task in the set was initiated first, followed by the 3rd task, then the 1st task, and so forth. If an individual does not initiate a particular task, then the corresponding position in the vector sequence is treated as missing based on the sequences that were performed by others in the population.

E. Activity Interruptions

In the case of activities that involve waiting for an event (for example, waiting for the Heating Pad to warm up), interrupting the activity to finish other tasks is both efficient and is an indication that the participant is capable of generating more complex plans that interweave multiple activities. However, for activities that take a short time to complete such as Change and Bus/Map, participants will likely complete the task without interruptions. To capture differences in interruptions on various activities, we define activity interruption features based on all DOT subtasks. The *activity interruption* feature monitors whether activities are in fact interrupted and interwoven. For long activities, such interruptions may indicate that the participant is able to generate a complex and efficient solution to the DOT.

The set of extracted features is summarized and categorized in Table 3. We hypothesize that these smart home features will allow us to provide automated task quality scores that correlate with task scores obtained by direct observation.

V. AUTOMATED SCORING

DOT task accuracy and task sequencing scores are derived from direct observation of participant's task performance. We used machine learning techniques to identify correlation between our automated feature set based on smart home sensor data and the direct observation scores. Here we describe two approaches to automated scoring: using supervised learning and using unsupervised learning techniques.

A. Supervised Scoring Models

We can formulate the automated scoring problem as a supervised learning problem in which a machine learning algorithm learns a function that maps the sensor-derived features to the direct observation scores. We use a support vector machine (SVM) with sequential machine optimization and bootstrap aggregation or bagging to learn the mapping. Support vector machines identify class boundaries that maximize the size of the gap between the boundary and data points. The booststrap aggregation improves performance of an ensemble learning algorithm by training the base classifiers on randomly-sampled data from the training set. The learner averages individual numeric predictions to combine the base classifier predictions and generates an output for each data point that corresponds to the highest-probability label. We use both supervised regression and classification algorithms in our supervised scoring models.

TABLE 3
DOT FEATURE SET

Feature Set	Feature Type
DOT features	Duration, sensor counts, sensor events, activity completeness
Interruption features	Number of activity interruptions
Sequencing features	Sequence vector
Parallelism feature	Pindex

B. Unsupervised Scoring Models

A score that is generated by a supervised learning algorithm predicts the quality of an activity in a way that emulates human-assigned scores. In contrast, unsupervised techniques use characteristics of the data itself to identify natural boundaries between activity performance classes. Here we derive unsupervised scores using a dimensionality reduction technique. Dimensionality reduction techniques reduce a high-dimensional dataset to one with a lower dimension. We use this to reduce the feature set to a single numeric score.

While we use Principal Component Analysis (PCA) to reduce the dimension, many reduction techniques would be appropriate for this task [21]. PCA is a linear dimensionality reduction technique that converts sets of features in a high-dimensional space to linearly uncorrelated variables, called principal components, in a lower dimension such that the first principal component has the largest possible variance, the second principal component has the second largest variance, and so forth. After reducing the dimension, we use min-max normalization to convert the variables to a uniform range.

C. Cognitive Assessment Models

In our final step, we evaluate the use of smart home techniques to automate the cognitive health assessment of participants based on sensor-based features that describe their activity performance. We map each participant to one of the three cognitive groups: Dementia (D), Mild Cognitive Impairment (MCI), or Cognitively Healthy (CH). To accomplish this, we extract the same sensor-based activity features that were used for the earlier experiments, as explained in Section 4. We obtain ground truth cognitive health labels for each participant from a battery of standardized and experimental neuropsychological tests that were administered in a clinical setting. These standard neuropsychological tests are conducted on a separate day from the smart home testing sessions. We then train machine learning algorithms to learn a mapping from the sensor-based activity features to the cognitive health label (CH, MCI or D).

VI. EVALUATION

Our goal is to design smart home technologies that assist with automated assessment of task quality and of cognitive health. Here we evaluate our approaches using data collected on a smart home testbed. We evaluate the two main tasks separately. To evaluate the ability to automate assessment of task quality, we compare scores generated from our smart home algorithm with direct observation scores generated from trained neuropsychologists. To evaluate the ability to automate assessment of cognitive health, we compare diagnoses generated from our machine learning-based smart home algorithms with diagnoses based on clinical tests.

We perform four experiments to evaluate our smart home-based task quality assessment algorithms. First, we measure the correlation between subsets of our smart home sensor features and direct observation scores (Section A.1). Second, we measure the correlation between the entire set of sensor features and direct observation scores (Second A.2). Third, we assess how well a support vector machine learning algorithm correctly classifies task quality, using the direct observation scores as ground truth labels (Section A.3).Finally, we determine how well the unsupervised learning-derived score correlates with direct observation scores (Second B.1).

In addition, we evaluate learning algorithms using different participant groups that we construct based on their cognitive diagnosis (D,MCI,CH) and number of subtasks that they complete. Since the number of cognitively healthy participants is large, we further divide them to older adults (MiddleAge, YoungOld, and OldOld) and younger adults (YoungYoung). These sample groups have different heterogeneity. We refer to a sample group as heterogeneous if it contains examples of both well-conducted and poorly conducted activities. Training set containing instances of cognitively healthy individuals who commit fewer mistakes tend to be less heterogeneous as compared to training set containing instances of both cognitively healthy individuals and individuals with MCI who often commit more mistakes. Similarly, individuals who complete fewer subtasks normally commit more mistakes than individuals who complete higher number of subtasks. By training learning algorithm using these sample subsets, we can understand how the heterogeneity impacts the performance of the learning algorithms and helps us to understand the features of these different groups.

We next evaluate the ability of our learning algorithm to map smart home activity sensor features to a cognitive health diagnosis. We train learning algorithms using smart home data and the cognitive health assessments provided by trained clinicians (Section C) and evaluate them using two metrics: the Area Under the ROC curve (AUC) and the F-score.

ROC curves assess the predictive behavior of a learning algorithm independent of error cost and class distribution. We plot false positives vs. true positive at various threshold settings to obtain a ROC curve. The area under the ROC curve (AUC) provides a measure that evaluates the performance of the learning algorithm independent of error cost and class distribution [47]. Similarly, F-score is the harmonic mean of the precision and recall and is defined as [47]:

$$F - score = 2 \times \frac{precision \times recall}{precision + recall}$$

A. *Evaluation of Supervised Scoring Models*

In this subsection, we evaluate our supervised scoring models.

A.1 Feature Subset Correlation

First, we consider alternative feature subsets as summarized in Tables 4 and 5. For each subset, we generate the correlation coefficient between the feature values derived from smart home sensor data and the experimenter direct observation scores (the accuracy score and sequencing score). In addition, we also analyze varying subsets of participants. Specifically, we consider subgroups of participants corresponding to the individuals with dementia (sample 1), individuals with MCI (sample 2), older adults (sample 3), and younger adults (sample 4). The objective of the experiment is to identify the correlation that exists between smart home task feature subsets for each participant and the activity quality score for the participant provided by trained clinicians and based on direct observation of the activity.

From Tables 4 and 5, we see that correlations between most of the feature subsets and direct observation accuracy / sequencing scores are statistically significant. We find that the correlation between the smart home features and the observation-based accuracy score is stronger than the correlation with observation-based sequencing scores. A possible reason for this is that the task *accuracy score* quantifies the correctness and completeness of the eight DOT subtasks, which reflects the same type of information captured by smart home features. In contrast, the *sequencing score* quantifies how the DOT subtasks were sequenced, which is not as extensively captured by smart home features.

We find that feature subsets correlate differently with different training sample subsets. For instance, in Table 4 DOT features have stronger correlation with task accuracy score but parallelism feature has weak correlation when we train learning algorithms with cognitively healthy group (column {4}). This indicates that a learning algorithm can better predict task accuracy with DOT features than parallelism features when training set contains examples of cognitively healthy individuals. Similarly, in Table 4 we see that the parallelism features correlates higher when sample subsets of training data contain individuals with MCI and younger adults (column {2,4}) but does not when it contains cognitively healthy individuals(column {4}) indicating that parallelism features can better represent differences between younger adults and MCI. Thus, we see that predictive power of a feature set depends on participant groups.

In addition, we visualize the relationship between selected feature types and the direct observation scores. Figure 6 plots the order in which subtasks were initiated within the DOT. As the figure shows, most participants placed Bus Map first in their sequence and almost all participants initiated the Exit subtask last. There is a fairly consistent choice of ordering among the subtasks for all participants, with the greatest variation occurring in positions 3, 6, and 7 of the sequence. From this figure, we conclude that task sequencing plays an important role in such a complex activity as the DOT and should be analyzed as a part of overall task quality.

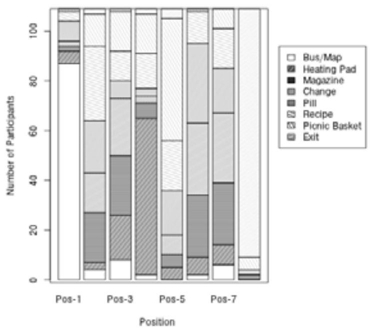

Fig. 6. DOT subtask order for the participants who completed all 8 subtasks. The x axis represents the subtask sequence position (1..8). The y axis represents the number of participants. Each bar corresponds to the number of participants that put a particular subtask in the given position of the subtask sequence order.

In a separate step, we plot the relationship between Pindex (the parallelism feature) and the direct observation scores. As shown in the left plot in Figure 7, Pindex consistently increases with accuracy score. The figure also shows a relationship between Pindex and the sequence score, although it is not as distinct. We note that when a participant initiates but does not complete subtasks their task quality degrades which increases their Pindex score. Correspondingly, as mentioned in Table 1, their accuracy score increases as well.

A.2 Combined Feature Correlation

In this experiment, we use the SMO-based support vector machine algorithm and bootstrap aggregation to learn a regression model that finds a fit between the combined set of feature values and the accuracy and the sequencing direct observation score. There are two objectives of this experiment. The first objective is to evaluate the correlation between the smart home DOT features and direct observation scores (accuracy and sequencing scores). The second objective is to study how the correlations between the smart home features and direct observation scores vary as different subsets of participants are considered. We first analyze the relationship for separate participant groups based on how many subtasks they completed then we look at the relationship for the participant groups based on their cognitive diagnosis. The results are summarized in Tables 6 and 7. In each table, the first row shows the correlation between the entire participant subgroup and the direct observation scores.

We find that the correlation depends on the heterogeneity in the samples. For example, the strongest correlation is found when examining the population subgroup that contains both MCI and cognitively healthy younger adults. On the other hand, the weakest correlation exists when considering only cognitively healthy individuals.

We find that the correlation decreases as participant subgroups that completed all activities are included. This is due to the variation that is present in the data for these individuals. In other words, having a large number of incorrect and inefficient tasks helps the learning algorithm to make a better prediction. The variations in the samples of cognitively healthy individuals who completed all activities are relatively low.

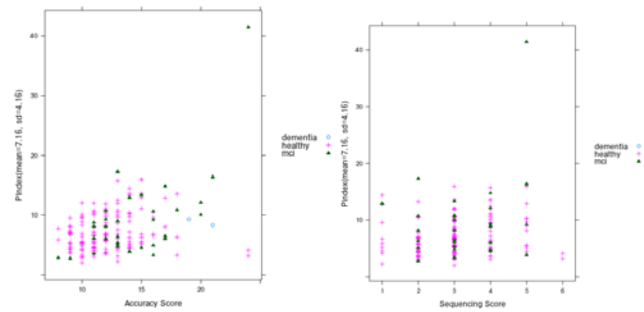

Fig. 7. Scatter plot of Pindex vs. Accuracy Score (left) and Pindex vs. Sequencing Score (right) grouped by cognitive diagnosis. The point in the upper right represents a participant who had difficulty completing any of the DOT subtasks.

We also find that the correlation is consistently stronger for the accuracy score than the sequencing score. This is because the accuracy score takes into account the mistakes that an individual makes in a subtask while the sequencing score only considers how a participant initiated an activity.

When we examine the correlation between the combined set of features and the direct observation scores for the entire population, we see that the coefficient is fairly high ($r=0.79$ for the accuracy score, $p<0.005$). This result indicates that automatically-derived feature values generated from smart home data do provide valuable information that can be used to assess task quality and that the quality score is fairly consistent with those obtained through direct observation.

A.3 Supervised Classification of Task Quality

In this experiment, we train multiple machine learning models to classify task quality score. Unlike the regression approach we considered in the earlier experiments, here we evaluate the ability of classification algorithms to predict task quality. We choose the observation-based accuracy score as our basis of comparison with automated scores because the correlation coefficients between features derived from sensor data with the accuracy score were consistently higher than the correlation between features from the sensor data and the sequencing score. We divide the scores into two classes using equal-frequency binning. Table 8 shows the results of the experiments when all samples are included. All results are generated using leave one out cross validation. The machine learning models that are tested include an SMO-based support vector machine, a neural network, and a naïve Bayes classifier. We see that learning algorithms are indeed effective at classifying task quality based on direct observation scores.

TABLE 4. CORRELATIONS BETWEEN FEATURE SUBSETS, PARTICIPANT GROUPS, AND THE ACCURACY DIRECT OBSERVATION SCORE. SAMPLES ARE: 1=DEMENTIA, 2=MCI, 3=OLDER ADULT, 4=YOUNGER ADULT

Correlation coefficient (r)							
Participant sample	{1,2,3,4}	{2,3,4}	{3,4}	{4}	{2,4}	{2}	{2,3}
Sample Size	179	177	145	37	69	32	140
DOT features	0.58** ǂ	0.57** ǂ	0.57** ǂ	0.52** ǂ	0.54** ǂ	0.44*	0.55** ǂ
Interruption features	0.31** ǂ	0.32* ǂ	0.25*	0.21	0.27	0.40	0.36*
Sequencing features	0.76** ǂ	0.72** ǂ	0.64** ǂ	0.36*	0.78** ǂ	0.79** ǂ	0.68** ǂ
Parallelism feature	0.39** ǂ	0.39** ǂ	0.18*	0.11	0.59** ǂ	0.58** ǂ	0.39** ǂ

*p<0.05, **p<0.005, ǂp<0.05 with Bonferroni correction for n sample groups

TABLE 5. CORRELATIONS BETWEEN FEATURE SUBSETS, PARTICIPANT GROUPS, AND THE SEQUENCING DIRECT OBSERVATION SCORE. SAMPLES ARE: 1=DEMENTIA, 2=MCI, 3=OLDER ADULT, 4=YOUNGER ADULT

Correlation coefficient (r)							
Participant sample	{1,2,3,4}	{2,3,4}	{3,4}	{4}	{2,4}	{2}	{2,3}
Sample Size	179	177	145	37	69	32	140
DOT features	0.10	0.01	0.21	0.21	-0.01	-0.27	-0.08
Interruption features	0.43** ǂ	0.42** ǂ	0.45** ǂ	0.47** ǂ	0.28*	0.22	0.34**
Sequencing features	0.46** ǂ	0.42** ǂ	0.50** ǂ	0.20	0.30*	-0.12	0.38** ǂ
Parallelism feature	0.12	0.13	0.03	0.32	0.29*	0.01	0.02

*p<0.05, **p<0.005, ǂp<0.05 with Bonferroni correction for n sample groups

B. Evaluation of Unsupervised Scoring Models

In this section, we evaluate unsupervised scoring models explained Section V.B.

B.1 Unsupervised Score Correlation

In our next experiment, we analyze the correlation between unsupervised learning model-based generation of a sensor-derived score using Principal Component Analysis and the direction observation-based accuracy score and sequencing score. The objective of this experiment is to test the performance of unsupervised learning models in predicting DOT activity quality scores and determine if the performance of an unsupervised algorithm is comparable to that of a supervised learning algorithm. We first analyze the relationship for separate participant groups based on how many subtasks they completed then we look at the relationship for the participant groups based on their cognitive diagnosis. The results are summarized in Tables 9 and 10. Figure 8 shows the plot of the PCA score that is obtained by reducing the feature space to a single dimension as a function of the accuracy and sequencing scores.

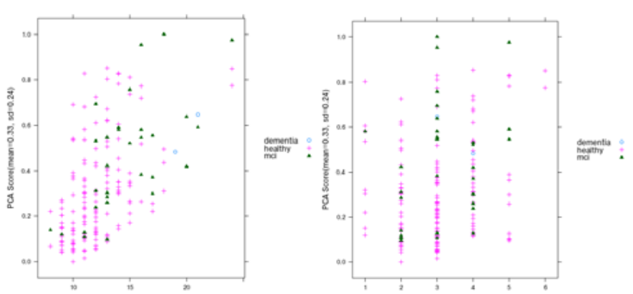

Fig. 8. PCA score vs. Accuracy score (left) and PCA score vs. Sequencing score (right).

Similar to previous observations, we find that the correlation depends on the heterogeneity in the samples. For example, the strongest correlation is found when examining the population subgroup that contains both MCI and cognitively healthy younger adults. The correlation coefficient between the unsupervised score and the direct observation accuracy score is 0.57 (p<0.005). This indicates that a fairly strong positive correlation exists between the automated scores and experimenter-generated scores of task quality. Furthermore, this value is similar to the values generated for the support vector machine model, which indicates that task quality can be computed directly using smart home sensor data without relying on training from human-provided scores.

C. Evaluation of Cognitive Assessment Models

We now turn our attention to evaluating the cognitive assessment models.

C.1 Cognitive Health Classification

The second goal of this project is to design a machine learning approach to automate cognitive health assessment based on smart home features. For this study, we map each participant to one of three labels: Cognitively Healthy (CH), Mild Cognitive Impairment (MCI), or Dementia (D). We use labels provided by clinical testing to train the learning algorithm. Note that this data is based on a battery of standardized and experimental neuropsychology tests administered in a laboratory setting and not on the smart home data. We initially handle the assessment as a set of binary classification problems.

One challenge that we face in learning a discriminative model between these three classes is that there is a tremendous class imbalance. While there are 145 cognitively healthy individuals, there are 32 individuals with MCI and only 2 participants with dementia. Part of this imbalance is because many dementia participants had such difficulty completing basic everyday tasks independently that they were not able to initiate the DOT. Class imbalance typically adversely affects classification performance because machine-learning models tend to label the points with the majority class label. To address this issue, we use cost sensitive versions of machine learning algorithms for each of the base classifiers. A cost sensitive classifier assigns misclassification costs separately for individual class labels and reweights the samples during training according to this cost. This allows the classifier to achieve overall strong performance even when the training points are not evenly divided among the alternative classes [40], as is the case with this dataset.

TABLE 6
CORRELATIONS BASED ON NUMBER OF SUBTASKS THAT ARE COMPLETED (*P<0.05, **P<0.005, ǂP<0.05 WITH BONFERRONI CORRECTION FOR THE SAMPLE GROUPS)

#Completed subtasks	Sample size (n)	Accuracy score	Sequencing score
2	179	0.79** ǂ	0.45** ǂ
3	174	0.77** ǂ	0.36** ǂ
4	172	0.76** ǂ	0.41** ǂ
5	167	0.75** ǂ	0.37** ǂ
6	154	0.65** ǂ	0.43** ǂ
7	137	0.57** ǂ	0.48** ǂ
8	83	0.43** ǂ	0.49** ǂ

TABLE 7.
CORRELATIONS BASED ONCOGNITIVE DIAGNOSIS (*P<0.05, *P<0.005, ǂP<0.05 WITH BONFERRONI CORRECTION FOR THE SAMPLE GROUPS)

Cognitive Diagnosis	Sample size (n)	Accuracy score	Sequencing score
{1,2,3,4}	179	0.79** ǂ	0.45** ǂ
{2,3,4}	177	0.80** ǂ	0.43** ǂ
{3,4}	145	0.75** ǂ	0.57** ǂ
{4}	37	0.70** ǂ	0.41** ǂ
{2,4}	69	0.81** ǂ	0.27*
{2}	32	0.75** ǂ	-0.09
{2,3}	140	0.78** ǂ	0.34** ǂ

TABLE 8
PERFORMANCE OF THE MACHINE LEARNING CLASSIFIERS ON THE SUPERVISED CLASSIFICATION OF TASK QUALITY

Learning algorithm	Accuracy	F-score		AUC
		Class A	Class B	
Support vector machine	80.45	0.84	0.76	0.85
Neural Network	79.33	0.82	0.74	0.85
Naïve Bayes classifier	82.13	0.85	0.78	0.88

TABLE 9
CORRELATIONS BASED ON NUMBER OF SUBTASKS THAT ARE COMPLETED USING PCA (*P<0.05, **P<0.005, ǂP<0.05 WITH BONFERRONI CORRECTION FOR THE SAMPLE GROUPS)

#Completed subtasks	Sample size (n)	Accuracy score	Sequencing score
2	179	0.57** ǂ	0.23** ǂ
3	174	0.46** ǂ	0.14
4	172	0.45** ǂ	0.13
5	167	0.50** ǂ	0.13
6	154	0.48** ǂ	0.13
7	137	0.47* ǂ	0.10
8	83	0.43** ǂ	0.10

In our first experiment, we train a machine learning algorithm to label Cognitively Healthy and MCI participants. We use PCA to reduce the dimensionality of the feature vector and train a cost-sensitive version of a support vector machine. We compare this with an alternative approach in which we handled the class imbalance by under-sampling the majority class. The results of this experiment are summarized in Table 11. To compare automated diagnosis based on smart home features with diagnosis based on direction observation features, we also train a learning algorithm to map direct observation-based scores to cognitive health diagnosis labels. The AUC value for this mapping is 0.68 in the best case (using naïve Bayes and under sampling). The predictive performance overall is not as strong as we would like to see for this case, in part because performance of Cognitively Healthy and MCI participants is actually quite similar on familiar activities such as those used in the DOT. The individuals in these two groups do have quite a bit of overlap in functional performance as is evident in Figure 7. MCI is considered as a transition from healthy aging to dementia so there is a wide range of cognitive difficulties in this class.

TABLE 10
CORRELATIONS BASED ON COGNITIVE DIAGNOISIS THAT ARE COMPLETED USING PCA (*P<0.05, **P<0.005, ǂP<0.05 WITH BONFERRONI CORRECTION FOR THE SAMPLE GROUPS)

#Cognitive Diagnosis	Sample size (n)	Accuracy score	Sequencing score
{1,2,3,4}	179	0.57** ǂ	0.23** ǂ
{2,3,4}	177	0.56** ǂ	0.23** ǂ
{3,4}	145	0.44** ǂ	0.17
{4}	37	0.06	0.47** ǂ
{2,4}	69	0.77** ǂ	0.38*
{2}	32	0.79** ǂ	0.32*
{2,3}	140	0.51** ǂ	0.17*

In a second experiment, we compare the Cognitively Healthy group with the Dementia group. We under-sampled the Cognitively Healthy class so that the ratio of Cognitively Healthy to Dementia would be 2:1. During testing we ensured that one Dementia participant would be used each time for training and the other would be used for testing. The results of this experiment are summarized in Table 12. These two groups are much easier to distinguish, which is to be expected. Note that out of 16 dementia participants that were recruited, 14 participants did not complete the DOT. One approach would be to represent the sensor features as missing for these participants and note the number of tasks that were completed as 0. In this case, the classification performance of the Dementia group would be almost 100% accuracy. These experiments provide evidence that the learning algorithm can provide an indication of the cognitive health of an individual based on activity performance.

VII. Observations

Researchers have hoped for quite a while that ubiquitous computing technologies could be used to support health monitoring and aging in place. This study provides an indication that with smart home sensor data and machine learning algorithms it is possible to automatically predict the quality of daily activities. In contrast, to date assessment of cognitive and physical health is traditionally performed outside the home or requires direct observation and analysis from trained professionals.

TABLE 11
PERFORMANCE OF THE MACHINE LEARNING CLASSIFIERS ON THE SUPERVISED CLASSIFICATION OF COGNITIVE HEALTH (MCI/ COGNITIVELY HEALTHY)

Learning algorithm	F-score		AUC
	Class A	Class B	
PCA + SMO with Cost Sensitive Learning	0.37	0.78	0.62
Under sampling of Majority Class + Bagged SMO	0.34	0.42	0.60

TABLE 12
PERFORMANCE OF THE MACHINE LEARNING CLASSIFIERS ON THE SUPERVISED CLASSIFICATION OF COGNITIVE HEALTH(DEMENTIA/COGNITIVELY HEALTHY)

Learning algorithm	F-score		AUC
	Class A	Class B	
Under sampling + Bagged SMO	0.54	0.54	0.56
Missing Values+ SMO	0.93	0.99	0.94

One must carefully interpret the results that we have mentioned. We note that the correlation (r) between smart home features and task accuracy scores is statistically significant. We can more conservatively analyze the correlation coefficient using a coefficient of determination. We square the correlation coefficient to obtain the coefficient of determination. A coefficient of determination of 0.62 (r=0.79) means that the 62% of the variation in the dependent variable can be explained by the variation in the independent variable. Our current results show that our method explains nearly 62% variations in the direct observational scores. Unexplained variation can be attributed to limitations of sensor system infrastructures and algorithmic limitations.

The implication is that smart home technologies can provide valuable information to assess the qualities of daily activities and can be implemented in home with improvements in sensor technology and algorithm design. In other hand, predicting cognitive health based on the performance on activities of daily living is an active research area in neuropsychology and clinical research [34][35][44]. Thus, we believe that smart home based technologies can monitor activities of daily living and predict cognitive health of an individual. Our results indicate this as a possibility.

We observe from the experiments that the performance of automatic task quality prediction depends on the type of training samples that are provided. The learning algorithm offers more accurate predictions when the training samples contain heterogeneous data points of both well-conducted activities and poorly-conducted activities. The algorithm does not predict well when instances of only one type are included. We also find that different feature subsets correlate with different types of training examples, categorized by age or by cognitive diagnosis. We observe that sequencing features are less indicative when all of the participant samples are cognitively healthy, while parallel features are indicative when we include MCI and younger adult participant samples. We therefore conclude that researchers need to carefully define and extract appropriate features from sensor data to use in building an assessment model. In addition, for our study the baseline for performance is a direct observation score based on coders observation of task performance. Two coders, blinded to group assignment, independently assigned scores to participants based on specific criteria as they directly viewed the participants' task performance. We cannot, however, ignore the fact that there may be some error or bias in these direct observation scores. This error can be mitigated by increasing the number of clinicians scoring the activities or by automatically detecting and correcting for bias.

We observe from our study that some limited cognitive health assessment can be automated using smart home sensors and algorithms. There are limitations of the current approach that can be improved in future work. One limitation is the coarse granularity of the home-based sensors. While environment sensors face fewer practical issues of user acceptance, placement, and battery charge, our algorithms would benefit from more detailed data provided by wearable, smart phone, and object sensors. In addition, our data collection was limited because many participants with cognitive difficulties were not able to complete the activities. We can address this issue by increasing our sample of participants. We note that the complexity of DOT was necessary to capture differences in task performance between cognitively healthy and MCI participants, but additional tasks that are less complex but still involve multi-tasking can be devised for future studies as well.

The study described in this paper is a step toward our overall goal of performing cognitive health assessment in smart homes. The direct observation scores mentioned in this paper are based on a traditional form of assessment in which patients travel to a lab or doctor's office and are tested by trained clinicians. In contrast, smart home systems continuously monitor individuals in their natural environment and provide ecologically-valid feedback on their everyday functioning. Clinicians and caregivers can use this information to make informed decisions about patient care.

There are a few limitations of our experimental methodology that need to be addressed. Our assessment technique relies on participants completing scripted activities in a single smart home setting. These types of methods are argued to be ecologically valid [7] but there is still the possibility that participants perform activities in an unnatural manner due to the unfamiliar environment, the scripted manner of the activity, or the awareness of being monitored. In addition, we currently use direct observation scores and clinician-based cognitive diagnosis as ground truth labels to train our learning models. Instead of using these ground truth labels, we would like to learn models based on differences in natural activity performance between individuals who are known to be cognitive healthy and those who are known to have cognitive difficulties. Finally, some of the features that we derive rely on human annotation of sensor data. Data annotation is a time-consuming process and human annotators can introduce errors. We can avoid this annotation step by using activity recognition algorithms that can recognize interleaved and parallel activities as well as activity steps.

There are additional issues to consider when implementing assessment systems in homes. First, we have to be careful when comparing activity performance since there are many unique ways to perform activities. Clustering algorithms can be used to group individuals together who have similar lifestyles and ways of performing activities and perform comparative assessment for these subgroups. Alternatively, we can use an individual as their own baseline and look for changes in activity performance over time that might indicate changes in cognitive health. Second, we need to carefully consider which activities to recognize, track, and use for assessment in everyday home environments. While sleep, exercise, and social interactions are common activities to monitor, other complex activity groups that involve multi-tasking are also useful for performing automated assessment. We will pursue these directions on our ongoing work.

VIII. Conclusions and Future Work

In this work, we showed that machine-learning algorithms can be designed to perform automated assessment of task quality based on smart home sensor data that is collected during task performance. Our preliminary results indicate that smart homes and ubiquitous computing technologies can be useful for monitoring complex everyday functions and to automate assessment of daily activities. This capability is valuable for monitoring the well-being of individuals in their own environments.

In our current work, we have focused on one complex activity, the DOT. We believe that the general approach is extensible to monitor a variety of activities, particularly for analysis across a population of individuals. In future work, we want to consider using a person as their own performance baseline and automate longitudinal assessment of well being and change in functional independence by analyzing sensor data collected in everyday environments over time.

Acknowledgements

The authors thank Selina Akhter, Thomas Cowger, Matt Kovacs, and Sue Nelson for their help in annotating the sensor data. We also thank Alyssa Hulbert, Carolyn Parsey, Chad Sanders, Courtney McAllister, Joyce Tam, Adri Seelye and Jennifer Walker for their assistance in collecting the data.

References

[1] M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox, and C.H. Phelps, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging Alzheimer's Association workshops on diagnostic guidelines for Alzheimer's disease," *Alzheimer's and Dementia: The Journal of the Alzheimer's Association*, vol. 7, pp. 270-279, 2011.

[2] S. Allin and E. Eckel, "Machine Perception for Occupational Therapy:Toward Prediction of Post-Stroke Functional Scores in the Home," in *Proc. Rehabilitation Engineering & Assistive Technology Society of North America (RESNA)*, 2006.

[3] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff, "A unified framework for gesture recognition and spatiotemporal gesture segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1685-1699, 2008.

[4] Alzheimer's Disease International, *World Alzheimer Report*. www.alz.co.uk/research/worldreport,2009.

[5] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, Psychiatric Press, 2000

[6] P. W. Burgess, "Strategy application disorder: The role of the frontal lobes in multitasking," *Psychological Research*, vol. 63, pp. 279-288, 2000.

[7] N. Chaytor, M. Schmitter-Edgecombe, and R. Burr, "Improving the ecological validity of executive functioning assessment," *Archives of Clinical Neuropsychology*, vol. 21, no. 3, pp. 217-227, 2006.

[8] D. J. Cook, "Learning setting-generalized activity models for smart spaces," *IEEE Intelligent Systems*, vol. 27, no. 1, pp. 32-38, 2012.

[9] D. J. Cook and S. K. Das, *Smart Environments: Technologies, Protocols, and Applications*, Wiley, 2004.

[10] D. J. Cook, and M. Schmitter-Edgecombe, "Assessing the quality of activities in a smart environment", *Methods of Information in Medicine*, vol. 48, no. 5, pp. 480-485, 2009.

[11] Department of Health. Speech by the Rt Hon Patricia Hewitt MP, Secretary of State for Health. in *Proc. Long-term Conditions Alliance Annual Conference*, 2007.

[12] X. Dong-Hui, A. S. Kurani, J. D Furst, and D. S. Raicu, "Run-length encoding for volumetric texture, " in *Proc. International Conference on Visualization Imaging and Image Processing*, pp. 68-73, 2004.

[13] S. T. Farias, D. Mungas, B. R. Reed, D. Harvey, D. Cahn-Weiner, and C. DeCarli, "MCI is associated with deficits in everyday functioning," *Alzheimer Disease and Associated Disorders*, vol. 20, no. 4, pp. 217-223, 2006.

[14] T. Giovannetti, B. M. Bettcher, L. Brennan, L., D. J. Libon, M. Burke,K. Duey,D. Wambach, "Characterization of everyday functioning in mild cognitive impairment: A direct assessment approach", Dementia and Geriatric Cognitive Disorders, vol. 25, pp. 359-365,2008.

[15] M. Hodges, N. Kirsch, M. Newman, and M. Pollack, "Automatic assessment of cognitive impairment through electronic observation of object usage," in *Proc. International Conference on Pervasive Computing*, pp. 192-209, 2010.

[16] T. L. Kasteren, G. Englebienne,, and B. J. Kröse,. "An activity monitoring system for elderly care using generative and discriminative models," *Personal and Ubiquitous Computing*, vol. 14, no.6, pp.489-498, 2010.

[17] E. Kim, A. Helal, and D. Cook, "Human activity recognition and pattern discovery," IEEE Pervasive Computing, vol. 9, pp. 48-53, 2010.

[18] J. R. Kwapisz, G. M. Weiss, and S.A. Moore, "Activity recognition using cell phone accelerometers," in *Proc. International Workshop on Knowledge Discovery from Sensor Data*, pp. 10-18, 2010.

[19] M.L. Lee and A.K. Dey, "Embedded assessment of aging adults: A concept validation with stake holders," in *Proc. International Conference on Pervasive Computing Technologies for Healthcare*, pp. 22-25, 2010.

[20] B. Logan, J. Healey, M. Philipose, E.M. Tapia, and S. Intille, "A long-term evaluation of sensing modalities for activity recognition," in *Proc. Conference on Ubiquitous Computing*, pp. 483-500, 2007.

[21] L. Maaten, E. Postma, and H. Herik, "Dimensionality reduction: A comparative review, " Tilburg University Technical Report, TiCC-TR 2009-005, vol. 10, pp. 1-35,2009.

[22] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on

different body positions, " in *Proc. International Conference on Body Area Sensor Networks*, pp. 113-116, 2006.

[23] P. Palmes, H. K. Pung, T. Gu, W. Xue, and S. Chen, "Object relevance weight pattern mining for activity recognition and segmentation," *Pervasive and Mobile Computing*, vol. 6, pp. 43-57, 2010.

[24] C. M. Parsey, and M. Schmitter-Edgecombe, "Direct observation of instrumental activities of daily living in normal aging, mild cognitive impairment, and dementia" Poster to be presented at the 41st annual meeting of the International Neuropsychological Society, Kona, Hawaii, February 2013.

[25] T. Pavan, C. Rama, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473-1488, 2008.

[26] M. Pavel, A. Adami, M. Morris, J. Lundell, T.L. Hayes, H. Jimison, and J.A. Kaye, "Mobility assessment using event-related responses," in *Proc. Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare*, pp. 71-74, 2006.

[27] R.C. Petersen, R. Doody, A. Kurz, R.C. Mohs, J.C. Morris, P.V. Rabins, and B. Winblad, "Current concepts in mild cognitive impairment," *Archives of Neurology*, vol. 58, pp. 1985-2992, 2001.

[28] R.C. Petersen and J.C. Morris, "Mild cognitive impairment as a clinical entity and treatment target," *Archives of Neurology*, vol. 62, pp. 1160-1163, 2005.

[29] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel, "Inferring activities from interactions with objects," *IEEE Pervasive Computing*, vol. 3, pp. 50-57, 2004.

[30] C. Phua, K. Sim, and J. Biswa, "Multiple people activity recognition using simple sensors," in *Proc. International Conference on Pervasive and Embedded Computing and Communication Systems*, pp. 313-318, 2011.

[31] M. Rabbi, S. Ali, T. Choudhury, and E. Berke, "Passive and in-situ assessment of mental and physical well-being using mobile sensors," in *Proc. International Conference on Ubiquitous Computing*, pp. 385-394, 2011.

[32] P. Rashidi, D. J. Cook, L. B. Holder, and M. Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 4, pp. 527-539, 2011.

[33] N. Ravi, P. Mysore, and M. Littman, "Activity recognition from accelerometer data," *In Proc. Conference on Innovative Applications of Artificial Intelligence*, pp. 1541-1546, 2005.

[34] M. Schmitter-Edgecombe, C. Parsey, and D. J. Cook, "Cognitive correlates of functional performance in older adults: Comparison of self-report, direct observation, and performance-based measures," *Journal of the International Neuropsychological Society*, vol. 17,no. 5, pp. 853-864, 2011.

[35] M. Schmitter-Edgecombe, C. McAlister, and A. Weakley, "Naturalistic assessment of everyday functioning in individuals with mild cognitive impairment: The Day Out Task. Neuropsychology", Neuropsychology, vol. 26, no.5, pp. 631-641, 2012.

[36] M. G. Sherod, H. R. Griffith, J. Copeland et al., "Neurocognitive predictors of financial capacity across the dementia spectrum: Normal aging, mild cognitive impairment, and alzheimer's disease", Journal of the International Neuropsychological Society, vol.15, no. 2, pp. 258-267,2009

[37] G. Singla, D. J. Cook, and M. Schmitter-Edgecombe, "Recognizing independent and joint activities among multiple residents in smart environments," *Ambient Intelligence and Humanized Computing Journal*, vol. 1, no. 1, pp. 57-63, 2010.

[38] A. J. Smola, and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing* , vol. 14, pp. 199-222, 2004.

[39] C. Spearman, "The proof and measurement of association between two things," *International Journal of Epidemiology*, vol. 39, no. 5, pp. 1137-1150, 2010.

[40] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, vol. 40, pp. 3358-3378, 2007.

[41] United Nations, World population ageing: 1950-2050, www.un.org/esa/population/publications/worldageing19502050/ , accessed May, 2012.

[42] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, "An activity monitoring system for elderly care using generative and discriminative models," *Personal and Ubiquitous Computing*, vol. 14, no. 6, pp. 489-498, 2010.

[43] G. Vincent and V. Velkoff, "The next four decades – the older population in the United States: 2010 to 2050," US Census Bureau, 2010.

[44] V. Wadley, O. Okonkwo, M. Crowe, D. Vance, J. Elgin, K. Ball, and C. Owsley, "Mild cognitive impairment and everyday function: An investigation of driving performance, " *Journal of Geriatric Psychiatry and Neurology*, vol. 22, no. 2, pp. 87-94, 2009.

[45] L. Wang, T. Gu, X. Tao, and J. Lu, "Sensor-based human activity recognition in a multi-user scenario," *LNCS*, vol. 5859, pp. 78-87, 2009.

[46] V. G. Wadley, O. Okonkwo, M. Crowe, and L. Ross-Meadows, "Mild cognitive impairment and everyday function: Evidence of reduced speed in performing instrumental activities of daily living", The American Journal of Geriatric Psychiatry, vol. 16 no.5, pp. 416-424, 2008.

[47] Ian H. Witten and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques", San Francisco, CA : Morgan Kaufmann Publishers Inc., 2005.

Prafulla N. Dawadi Prafulla Dawadi is currently a PhD student in the School of Electrical Enginering and Computer Science at Washington State University. He received his Bachelors in Computer Engineering from Advanced Engineering College, Tribhuwan University, Kathmandu, Nepal in 2007. His research interests include machine learning, data mining and smart environments.

Diane J. Cook Dr. Diane J. Cook is a Huie-Rogers Chair Professor in the School of Electrical Engineering and Computer Science at Washington State University. She received a B.S. degree in Math/Computer Science from Wheaton College in1985, a M.S. degree in Computer Science from the University of Illinois in 1987, and a Ph.D. degree in Computer Science from the University of Illinois in 1990. Her research interests include artificial intelligence, machine learning, and smart environments.

Maureen Schmitter-Edgecombe Dr. Maureen Schmitter-Edgecombe is a Professor in the Department of Psychology at Washington State University. Dr. Schmitter-Edgecombe received a B.S. from Bucknell University in 1988 and a M.S. and Ph.D. from the University of Memphis in 1991 and 1994, respectively. Dr. Schmitter-Edgecombe's research focuses on evaluating attention, memory, and executive functioning issues in neurological normal and clinical populations with the goal of designing and assessing rehabilitation techniques.