Active Learning with Uncertainty Sampling for Large Scale Activity Recognition in Smart Homes

Hande Alemdar*[1], T.L.M van Kasteren[2], and Cem Ersoy[1]

[1]Boğaziçi University, Department of Computer Engineering, Istanbul, Turkey
[2]Schibsted, Barcelona, Spain

Abstract

One of the major problems faced by automated human activity recognition systems is the scalability. Since the probabilistic models employed in activity recognition require labeled data sets for adapting themselves to different users and environments, redeploying these systems in different settings becomes a bottleneck. In order to handle this problem in a cost effective and user friendly way, uncertainty sampling based active learning method is proposed. With active learning, it is possible to reduce the annotation effort by selecting only the most informative data points for annotation. In this paper, three different measures of uncertainty have been used for selecting the most informative data points and their performance have been evaluated by using real world data sets. It has been shown that the annotation effort can be reduced by a factor of two to four, depending on the house and resident settings in an active learning setup.

1 Introduction

It is foreseen that in the near future, smart environments that interact with the people according to their specialized needs will be become an inseparable part of daily life. Hence, recognizing human activities in an automated manner is essential in many ambient intelligence applications such as smart homes and health monitoring and assistance applications [1]. In order to make such long term health monitoring systems sustainable, smart environments that recognize human activities automatically are needed [2, 3].

During the past decade, the advances in the sensor technology and wireless communication networks in terms of capacity increase, cost efficiency and power efficiency made it possible to use sensors for human activity recognition purposes.

*hande.alemdar@boun.edu.tr

These miniaturized sensors are soon to be deployed in large scale in many houses with highly varying sensor placements and layouts and produce vast amount of data. As the data supply increases, the demand for techniques to process such a huge amount of data in order to extract useful information in a reasonable amount of time also increases. In order to meet this demand, data-driven methods that are easily applicable to novel settings can be employed. In order to infer the human activities from the data collected from smart environments machine learning methods are needed, but those methods require annotated data sets to be trained on. Recording and annotating such data sets are costly since they require time and human effort. Although the annotated data sets are essential, they are hardly useful when recorded in laboratory settings following predefined scenarios since they do not reflect the natural human activities. Moreover, when automatic human activity monitoring systems are deployed on a world-wide scale for health care purposes, it is needed to fine tune the model behavior, which is characterized by a set of parameters, for each new house in order to accurately reflect the residents' activities for that specific house. However, in order to obtain adequate activity recognition accuracy in a new house, several weeks of annotated data from that specific house is needed. Instead of recording and annotating several weeks of data fully, an intelligent algorithm can be used to decide for which point in time it would be most informative to obtain annotation. Using the smart algorithm, the activity monitoring system can prompt the resident and ask which activity is currently being performed. This would minimize the need for annotation and maximize the usefulness of annotation.

All of the probabilistic models proposed for human activity recognition require labeled training data to learn the model parameters. Two problems limit the applicability of these models on a large scale with many different houses: (i) Differences in the layout of houses and the differences in the ways activities are performed by different people imply that a set of model parameters used for one house cannot be used directly in another house. (ii) The behavior of inhabitants changes over time, therefore parameters learned at one point in time may not accurately represent the behavior at a later point in time. Although both of these problems can be resolved by recording further annotated data, this solution is far from being practical and cost effective. Instead, novel learning methods that allow to deal with these problems cost effectively are proposed. This would allow the installation of activity recognition systems on a large scale with many different houses with different layouts and for a diverse population of inhabitants. This scalability provides a solution for dealing with the consequences of an aging population.

In order to decrease the annotation effort, a machine learning technique called *active learning* to select only the most informative data points for annotation has been used. In that way, the amount of training data needed is reduced and the annotation effort has been minimized. In this study, a framework for active learning that can be used with any probabilistic model has been proposed and the performance of the proposed method has been evaluated by conducting experiments on the multiple real world data sets. The results show that proposed active learning scheme

achieves a nearly equal recognition performance to the fully labeled data case by using only 7% of all data points and after a few days of data collection. Since fully annotation of the activities is marking the start and end times of an activity, the actual cost of annotation is not equal to the number of data points available. Hence, it has been shown that with the proposed method, the actual annotation cost is reduced by 30% to 75% while achieving the same recognition performance. Finally, a web-based annotation tool has been provided to show how the proposed active learning setup can be achieved with a user-friendly and efficient application. The study not only reports the best results reported so far but also provides an end-to-end solution to the scalability problem of activity recognition in smart environments. The applicability of a cost efficient and user-friendly method in future smart homes with automatic activity recognition capability has been demonstrated.

The paper is organized as follows. In Section 2, a brief literature review on active learning applications to activity recognition is given. In Section 3, the details of the model and active learning methods used are provided. Section 4 gives the details of the experiments with real world data. In Section 5, the web-based application for collecting the annotation labels are provided. Finally, Section 6 concludes the paper.

2 Related Work

The idea of using interaction-based ambient sensors for home automation in an intelligent way was first presented in the late 90s [4]. The studies that use those sensors for activity recognition purposes started in the early millennium. The GatorTech smart house was built for research on ambient assisted living [5]. The house contained several smart appliances equipped with sensors such as a smart refrigerator in order to monitor food usage. A similar project called AwareHome [6] used several ceiling mounted cameras and radio frequency identification (RFID) tags for localization purposes. In Universal Knowledgeable Architecture for Real-Life appliances (UKARI) project, the researchers developed a distributed service platform for managing the networked appliances in a home network service [7]. These projects are among the first examples of living laboratories and they aimed developing a proof of concept.

One of the major benefits of these smart home projects is that they offer automatic activity recognition capability that can be used in many applications such as health care monitoring, security services, and energy management [8, 9]. In terms of activity recognition purposes, one of the earliest studies is the House_n project by Tapia *et al.* [10]. They installed reed switches and piezoelectric switches on doors, windows, cabinets, drawers, microwave ovens, refrigerators, stoves, sinks, toilets, showers, light switches, lamps, some containers and electronic appliances in two different houses in order to detect more than 20 activities. Over the years, several other researchers also conducted similar studies in order to automatically infer the activities of daily living in smart environments [11, 12, 13, 14].

A recent literature survey of state-of-the-art AAL frameworks, systems and platforms to identify the essential aspects of AAL systems was provided in [15]. Their review revealed that only 12 projects out of many continued their projects beyond the pilot phase and deployed their solutions into the real world, either at care facilities or private homes. Their findings indicate that the scalability issues and the reusability of the knowledge obtained previously should be addressed in the following studies. As a remedy to these issues, active learning based solutions should be considered.

Active learning has been generally used in part of speech tagging problems in natural language processing [16, 17]. More recently, there are studies that study active learning in deep learning for various applications such as sentiment analysis [18] and handwritten digit recognition [19].

Active learning in activity recognition systems is studied by other researchers using mobile and wearable sensors [20] and video-based sensing, mostly. Truyen *et al.* [21] propose an active learning method for a video-based activity recognition system. They use generative and discriminative temporal probabilistic models for recognizing activities from video sequences. In [22], Hasan *et al.* consider the problem of updating the models continuously from streaming videos. In their framework, unlabeled new instances continuously arrive and they automatically select the most suitable features to improve the existing model incrementally using a combination of deep networks and active learning. They learn the features in an unsupervised manner using deep networks and use active learning to reduce the amount of manual labeling of classes. While it is natural to recognize human activities in public space using video cameras, their use inside the home are not very well received due to privacy issues. In this study, active learning with a different form of sensing mechanism having different data modality has been used. In that sense, this study complements the previous work that uses video-based sensing.

There are a number of query selection strategies in the literature which are summarized in [23]. In [24], Liu *et al.* use active learning with a decision tree model to classify the activities collected by a group of wearable sensors. In [25], a similar study is presented using classifiers like decision tree, joint boosting and Naive Bayes. In both studies, uncertainty based active learning methods are employed and active learning has been showed to work well. On the other hand, since human activities are temporal in nature it would be more desirable to use models that consider the temporal nature of human activities such as hidden Markov models (HMMs) and conditional random fields (CRF)s. In [26], the authors propose to use active learning for adapting to the changes in the layout of the living place. They use an entropy based measure to select the most informative instances and they evaluate the performance with controlled experiments in laboratory. They show that active learning only needs 20% of the new training data to achieve almost the same recall and precision performance after deployment changes. In this study, it has been shown that only 7% of all the data is needed in a start from scratch scenario rather than a change in the deployment.

Zhao *et al.* [27] address the quality issues in crowd-sourced annotation systems

for large-scale labeling. Their study shows that an approach using the raw annotations obtained from Mechanical Turk platform in an Support Vector Machine (SVM) classifier with standard margin criteria active learning fails due to noisy annotations. Instead, they propose a Bayesian framework together with a measure for selecting the most informative data points that uses both local and global measures. In this way, they improve the accuracy of active learning in noisy real-world conditions, yielding classifiers with accuracies closer to those trained using ground-truth data. In [28], a similar study is presented. Their system uses activity labels collected from crowd-sourced annotators to train an online activity recognition system. In order to handle the real-time training of activities, they merge the input from multiple annotators into a single ordered label set.

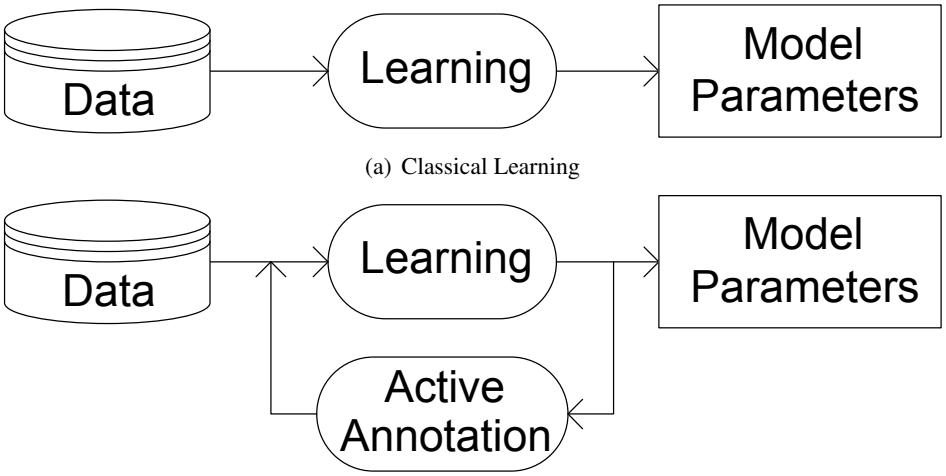
In [29], the authors present a non-probabilistic approach to activity recognition problem. They use data mining techniques such as frequent item set mining to sensor firings and cluster the data such that each cluster represents instances of the same activity. Then, the annotators only need to label each cluster as an activity as opposed to labeling all instances of all activities. After the associations between the clusters and activities are complete, the system can recognize future activities.

Bagaveyev and Cook [30] report promising results on CASAS smart-home data sets using a crowd-sourcing application for annotation. They experiment with two different active learning approaches: they use an expected entropy based method together with a logistic regression and secondly, they use a committee based active querying method with random forest classifiers. They state that an active learning solution should be able to adapt itself to the budget limitations, complexity of the algorithm and the required performance measure.

In this work, an adaptive solution has been proposed with an efficient and low time complexity classifier. Unlike most of the related work, it has not been assumed that the data is presegmented. For this reason, a temporal model has been used, since it suits the temporal nature of the human activities and handle the segmentation problem with probabilistic modelling. Three different measures of uncertainty have been used for selecting the most informative data points and their performance has been evaluated by using real world data sets. It has been shown that only 7% of all the data points are enough for getting the same performance with the fully labeled data, also it has been shown that the actual annotation effort can be reduced by a factor of two to four, depending on the setup.

3 Active Annotation

Active learning is a special form of machine learning in which the learning algorithm is able to interactively choose the data points to be labeled. The annotator or other information sources, then, provide the desired labels for the selected data points. Based on the newly labeled inputs, the learning algorithm iteratively continues to learning procedure until some criteria are met, i.e., the desired accuracy level or the budget limits. This scheme is especially useful for situations in which



(a) Classical Learning

(b) Active Learning

Figure 1: Learning frameworks.

the unlabeled data is abundant but obtaining the labels is expensive. Active learning is also referred to as optimal experimental design in statistics.

In the remaining of the section, a brief information about existing machine learning techniques that do not use active learning are provided. After that, proposed active learning framework is described. Finally, three measures that can be used in active learning for selecting the most informative data points are given.

In order to use a probabilistic model, a set of model parameters have to be learned. In Figure 1(a), the classical learning framework is depicted. The model parameters which are denoted by θ, can be learned using a supervised method which only uses the data whose labels are obtained through annotation.

In the proposed framework, only the labeled data points are used for obtaining the model parameters and the unlabeled data is disregarded. As depicted in Figure 1(b), the active learning algorithm iteratively

1. Learns new parameters using supervised learning

2. Selects the most informative data points according to the current model parameters and obtain their labels

More formally, let $x = \{x_1, x_2, ...x_T\}$ be the set of data points (i.e. data collected from the sensors), $y = \{y_1, y_2, .., y_T\}$ be the set of true labels (i.e. activity performed by the user). The *labeled data set* is $\mathcal{L} = \{x_i, y_i \mid x_i \in x, \quad y_i \in y, \quad 1 \leq i \leq T\}$. The *unlabeled data set* is $\mathcal{U} = \{x_i \mid x_i \notin \mathcal{L}, \quad 1 \leq i \leq N\}$. Typically more unlabeled data is available than labeled data, $N \gg T$. The union of these data sets is defined as $\mathcal{D} = \{\mathcal{L} \bigcup \mathcal{U}\}$ and the size of $\mathcal{D}$ is fixed.

At each iteration, data points are transferred from $\mathcal{U}$ to $\mathcal{L}$ by performing annotation. The size of $\mathcal{L}$, denoted by T, increases while the size of $\mathcal{U}$, denoted by N,

decreases. The data points that will be transferred from $\mathcal{U}$ to $\mathcal{L}$ are selected by the active learning method according to some informativeness measure. Uncertainty is used for assessing the most informative data points [31]. Probabilistic models need to calculate the probability distribution of the activities at each data point to perform inference. For many probabilistic models, there exist efficient algorithms to calculate these quantities, for example, the forward-backward algorithm is used for HMMs [32]. The forward-backward algorithm gives the probabilities for each activity at each time slice. While performing the inference, the model selects the activity that has the highest probability value for that time slice. The forward-backward algorithm is used to obtain P_θ, which is the probability distribution of each activity at each time slice under the current model parameters θ. After that, to select the most informative data point, x^*, three different methods are used.

1. *Least Confident Method* considers only the most probable class label and selects the instances having the lowest probability for the most likely label.

$$x^* = \arg\max_x (1 - P_\theta(\hat{y} \mid x)) \tag{1}$$

   where $\hat{y} = \arg\max_y P_\theta(y \mid x)$ is the class label with the highest probability according to the current model parameters θ.

2. *Margin Sampling* selects the instances that the difference between the most and the second most probable labels is minimum.

$$x^* = \arg\min_x (P_\theta(\hat{y_1} \mid x) - P_\theta(\hat{y_2} \mid x)) \tag{2}$$

   where $\hat{y_1}$ and $\hat{y_2}$ are the two most probable classes.

3. *Entropy based* method selects the instances that have the highest entropy values among all probable classifications.

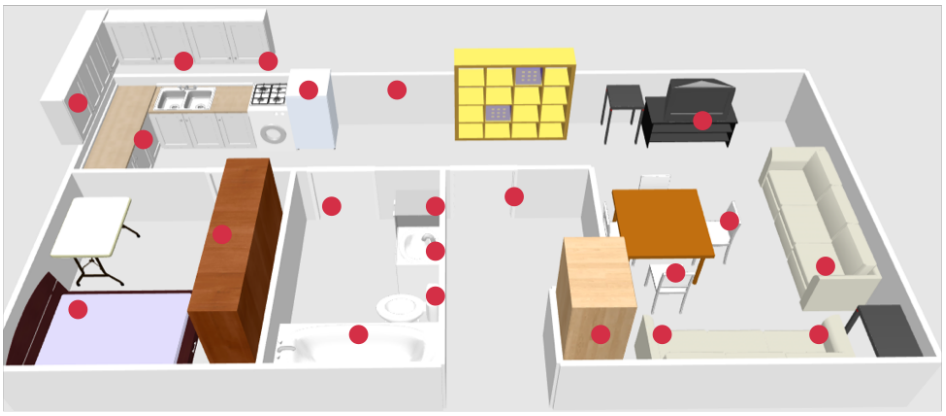
$$x^* = \arg\max_x -\sum_i (P_\theta(\hat{y_i} \mid x) log P_\theta(\hat{y_i} \mid x)) \tag{3}$$

4 Experiments

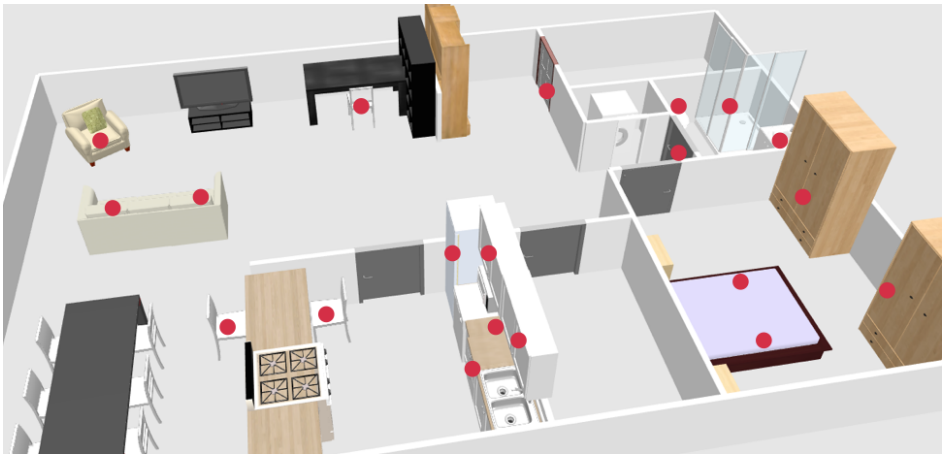
In the experiments, the effect of active learning for reducing the annotation effort in activity recognition is evaluated. The goal is to recognize the activities as accurate as possible while using the minimum amount of labeled data. Also, it is important not to disturb the annotator for a label that he/she possibly does not remember. Asking about the label of the activity that had been performed a month ago is not realistic. In this study, a daily querying approach has been used and its performance has been evaluated on real world data sets. The experiments aim to answer three questions: (i) Does active learning reduce the annotation effort?, (ii) What is the best uncertainty measure for selecting the most informative data points?, and (iii) What is the most suitable setup for the number of data points and for the number of iterations?

4.1 Data Sets

Publicly available ARAS human activity recognition data sets that are collected from two different real houses are used in the study. Each house was equipped with 20 interaction-based binary sensors of different types. A full month of information which contains both the sensor data and the activity labels for both residents was gathered from each house, each with two residents, resulting in a total of two months data. Overall, the number of recorded activities in House A and House B is 14 and 12, respectively. The details about the two houses (annotated as House A and B), the deployed systems, the residents and the collected data are given in [33]. The detailed layouts of Houses A and B along with the locations of the deployed sensors are presented in Figures 2(a) and 2(b), respectively.



(a) House A



(b) House B

Figure 2: House layouts and sensor deployments in ARAS data sets.

4.2 Experimental Setup

Markov models are widely used in the literature for modeling sequential data because they are well suited for handling the temporal dependencies. Since human activities are sequential in nature, Markov models have already proven to be useful for human activity recognition purposes. Linear Chain Conditional Random Fields (LCCRF) are also suited for modeling sequential data and they can be used in the proposed active learning scheme as with any other probabilistic model. However, due to the high run-time complexity required by LCCRFs, they are not well-suited for online settings. For that reason, HMM, which provide fast yet efficient learning algorithms that can complete in a few seconds, is preferred.

HMM is depicted in Figure 3. The hidden state at time t, denoted as y_t, correspond to the activities performed and the observations, x_t^i correspond to i^{th} sensor's value at time t. Each sensor modeled as an independent binary feature. The total number of sensors (features) is $N = 20$ for ARAS data sets. The total number of time steps is denoted as T. HMM is a generative model that has three factors in the joint probability distribution:

$$p(y_{1:T}, x_{1:T}) = p(y_1) \prod_{t=2}^{T} p(y_t \mid y_{t-1}) \prod_{t=1}^{T} p(x_t \mid y_t) \tag{4}$$

The initial state distribution $p(y_1)$ is a multinomial distribution; the transition distribution $p(y_t \mid y_{t-1})$ is represented as a collection of Q multinomial distributions (Q is the number of different activities); the observation distribution $p(x_t \mid y_t)$ is a multiplication of N independent Bernoulli distributions (N is the number of sensors).

$$\begin{aligned} p(x_t \mid y_t) &= \prod_{i=1}^{N} p(x_t^i \mid y_t) \\ p(x^i \mid y = j) &\sim Ber(\mu_{ij}) \end{aligned} \tag{5}$$

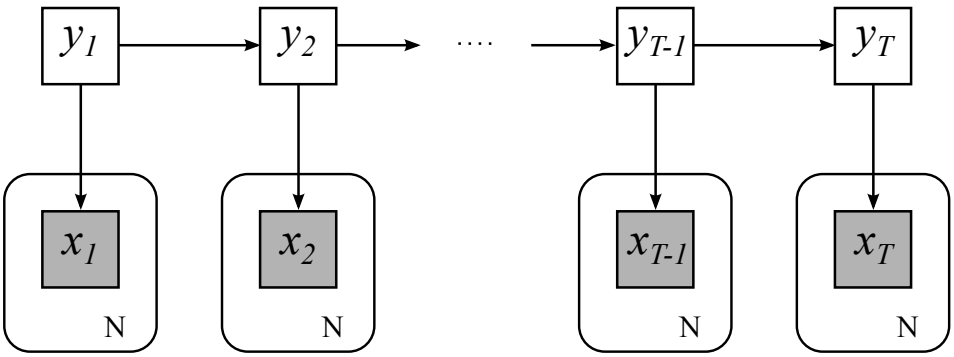


Figure 3: Hidden Markov model for activity recognition using N binary sensors.

A supervised approach with the maximum likelihood method for learning the parameters and the well-known Viterbi algorithm for inference is used. In order to prevent zero probabilities, Laplace smoothing is employed during parameter learning.

Sensor data is discretized in $\Delta t = 60sec$ intervals. Overall, there are $T = 1440$ data points for each day. For each sensor, the value 1 is used if the sensor has been fired at least once during the interval. For the ground truth labels used in training phase, the activity label that has the largest number of occurrences during that interval has been used. Leave-one-day-out cross validation is used in all experiments such that, one full day of data is held for testing and the remaining days are used for training. All days are used once for testing and the results are averaged over all folds. The training days are used in a sequential manner, that is, after a day's data has been processed, the algorithm moves to the following day and does not use the data of the previous day for obtaining labels. As stated previously, the algorithm iteratively learns new model parameters and selects the most informative points to be annotated. In the learning phase, all the data points whose labels are already obtained are used. However, data points to be annotated are only selected from the current day. In other words, in each iteration, the model parameters are learned with all the data obtained thus far. After that, according to the newly learned parameters, the data points to be annotated are selected from only the current day.

For measuring the performance, daily f-measure performances are used. For a multi-class classification problem, the metrics averaged over the number of activity classes are defined as follows:

$$Precision = \frac{1}{Q}\sum_{i=1}^{Q}\frac{TP_i}{TP_i + FP_i} \tag{6a}$$

$$Recall = \frac{1}{Q}\sum_{i=1}^{Q}\frac{TP_i}{TP_i + FN_i} \tag{6b}$$

$$F - measure = 2\frac{Precision.Recall}{Precision + Recall} \tag{6c}$$

where Q is the number of classes, TP_i is the number of true positive (TP) classifications for class i, FP_i is the number of false positive (FP) classifications for class i, and FN_i is the number of false negative (FN) classifications for class i.

With respect to the research questions to be answered, (i) a random selection approach together with uncertainty sampling is used to show the effect of the active selection, (ii) three different uncertainty measures are used to find the most suitable measure for selecting the most informative data points, and (iii) four different setups of active annotations are explored, namely, selecting

1. a single data point from each day in a single iteration,

2. ten data points from each day in a single iteration, resulting in ten data points from each day,

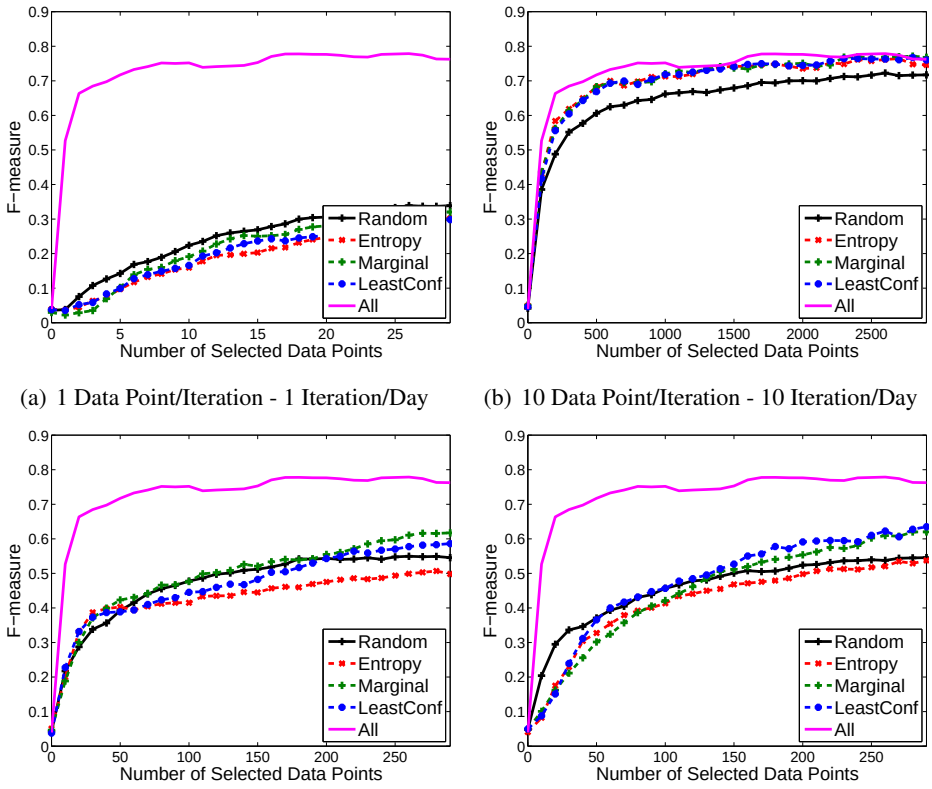
3. a single data point with ten iterations per day, resulting in ten data points from each day, and

4. ten data points in ten iterations per day, resulting in 100 data points from each day

4.3 Results

The results of the experiments for each house and for each resident are presented separately. For each case, the fully annotated data performance is included in the graphs in order to make realistic evaluations. The fully annotated performance graphs, drawn as solid magenta lines, indicate the scenario in which the whole 1440 data-points are selected from each day for annotation as opposed to the actively or randomly selected portions. The results for House A for Resident 1 is given in Figure 4. The results show that with a single data point from each day, the maximum achievable performance is severely degraded. With ten data points in ten iterations case, on the other hand, a highly comparable performance is observed with active learning. When the data points are randomly selected, it is not possible achieve the optimum performance. When the ten data points per day configurations are considered, similar performances with one iteration and ten iterations cases are observed. For these configurations, entropy based selection under-performs when compared to other selection methods.

In Figure 5, the results for House A, Resident 2 are shown. Similarly, the single point per day case yields a very low performance whereas the 100 points case reveals a significantly higher performance. Also, it is interesting to observe a higher performance than the fully annotated case. This can be attributed to the change in the resident's annotation behavior. The downward trend in the performance towards the end supports this argument. When full annotation is available, the observation model changes according to the annotator's overall average behavior immediately. When a difference in the way a specific activity is performed occurs, or a difference in the annotation behavior is observed, it is directly reflected on the performance. For example, consider *Watching TV* activity which is characterized by the use of remote control sensor and the pressure sensor on the couch. Consider a case where the user changed his favorite couch while *Watching TV* for a few occurrences. When full annotation is available, the observation model is updated right away so that the sensor firing probabilities for both the favorite couch and the new couch will be affected in the new model. If this change is temporary and the original favorite couch is being mostly used afterwards, then the change in the observation model is unnecessary and causes a degradation in performance. With an active learning scheme, if those data points are not selected for annotation, the discrepancy between the training and the test sets does not occur and no effect on the performance on the test sequence is observed. In other words, if no data points from *Watching TV* are selected during active annotation, the original observation model does not change. Although, in this case a higher performance



(a) 1 Data Point/Iteration - 1 Iteration/Day

(b) 10 Data Point/Iteration - 10 Iteration/Day

(c) 1 Data Point/Iteration - 10 Iteration/Day

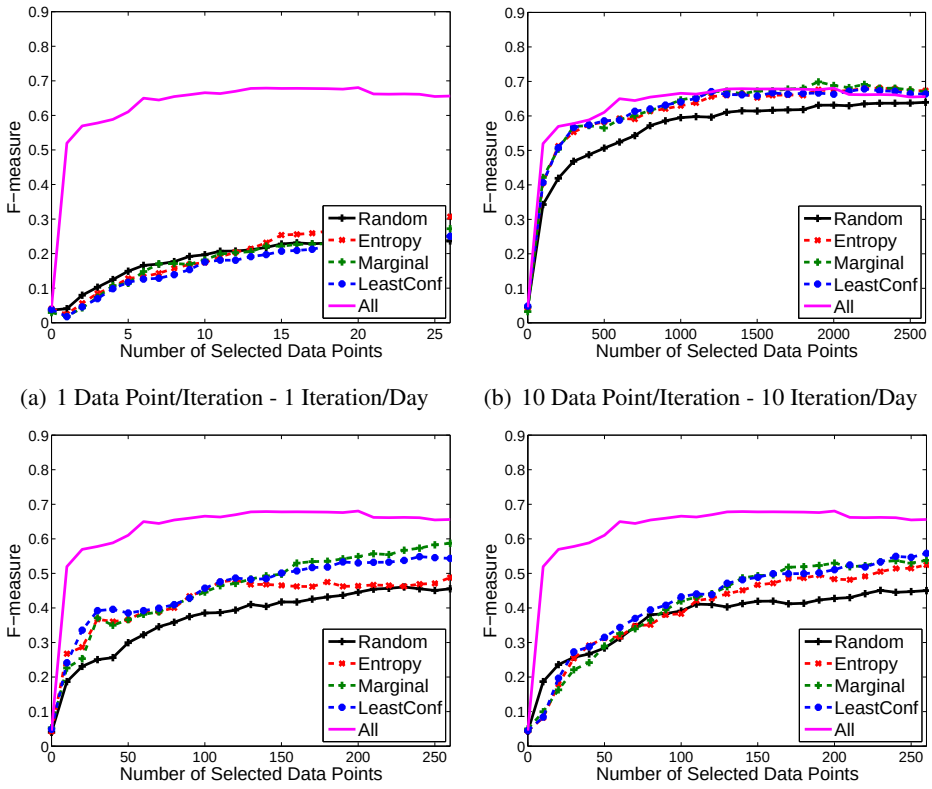
(d) 10 Data Point/Iteration - 1 Iteration/Day

Figure 4: Active learning experiment results for House A - Resident 1.

with active learning is obtained since the original activity behavior is resumed and the behavior change is temporary, it is important to note that this effect can also cause a degradation in the performance of active learning for other settings. It may cause some behavior changes to be understood later, as well. That is why, in order to capture the changes in behavior of the residents active learning systems should be used continuously rather than just at the beginning of new deployments.

The results for House B for the first resident is depicted in Figure 6. Most of the previous findings persist for this configuration as well but with a higher general performance increase with respect to the maximum achievable performance. With a 100 point selection per day, the performance converges to the maximum within five days. Also, the benefit of using uncertainty based measures over the random selection is more prominent in this house.

Finally, the second resident for House B results are given in Figure 7. Similar to the other resident's case for this house, the benefit of using active learning even with a low number of data points is prominent. With a single data point per day, the performance of marginal selection method is better than the other methods. For the other cases, there are not significant differences between the selection methods.



(a) 1 Data Point/Iteration - 1 Iteration/Day

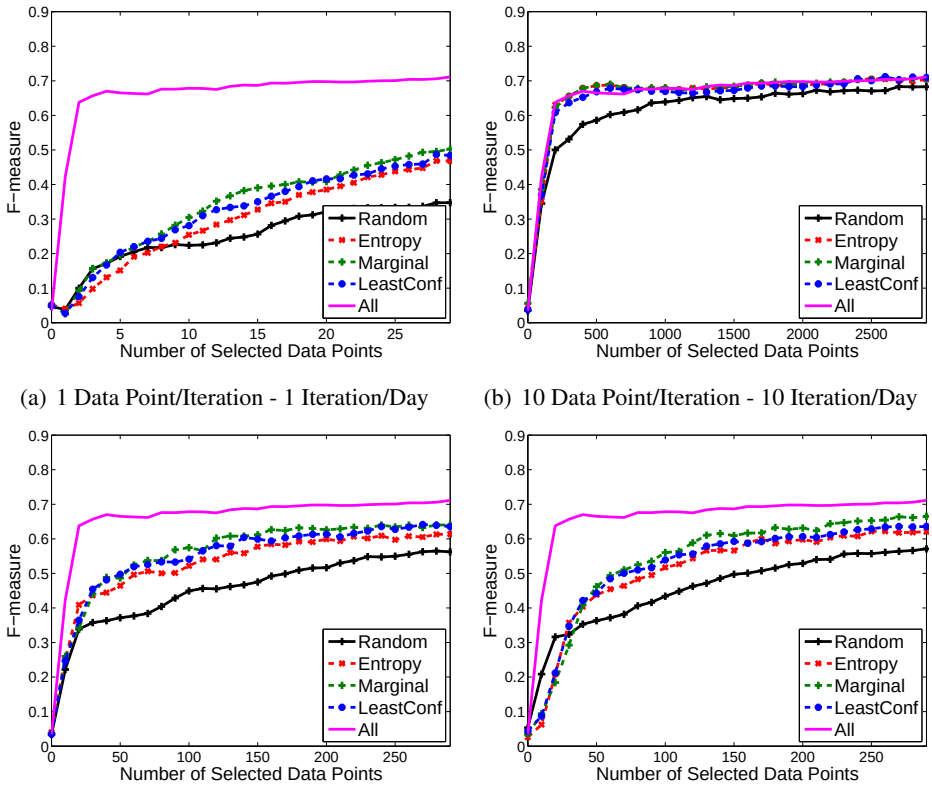
(b) 10 Data Point/Iteration - 10 Iteration/Day

(c) 1 Data Point/Iteration - 10 Iteration/Day

(d) 10 Data Point/Iteration - 1 Iteration/Day

Figure 5: Active learning experiment results for House A - Resident 2.

In the experiments, one-minute discretization is used, therefore, in each day there are 1440 data points. With the best performing setup that selects 100 points in each day, only 7% of all the available data points are used and fully annotated recognition performance is achieved only after a couple of days. The annotation effort in activity recognition is different than other machine learning problems such as image recognition or regression. Since the activities are continuous blocks, annotations are needed only for marking the start and the end of these activity blocks. Therefore, the total number of annotations required is not equal to the total number of data points. In order to take into account this fact, a separate analysis on the annotation cost reduction is provided. When the annotation effort of real time setup (i.e. at the time of data collection and whenever an activity starts or finishes) is compared against active annotation (i.e. after the day is completed in an offline fashion), again the offline active annotation is more preferable. During the data collection phase, the first and the second residents in House A made an average of 43 and 30 annotations per day, respectively. Similarly, for House B, the average number of annotations per day was 21 and 14 for resident 1 and 2, respectively. In



(a) 1 Data Point/Iteration - 1 Iteration/Day

(b) 10 Data Point/Iteration - 10 Iteration/Day

(c) 1 Data Point/Iteration - 10 Iteration/Day

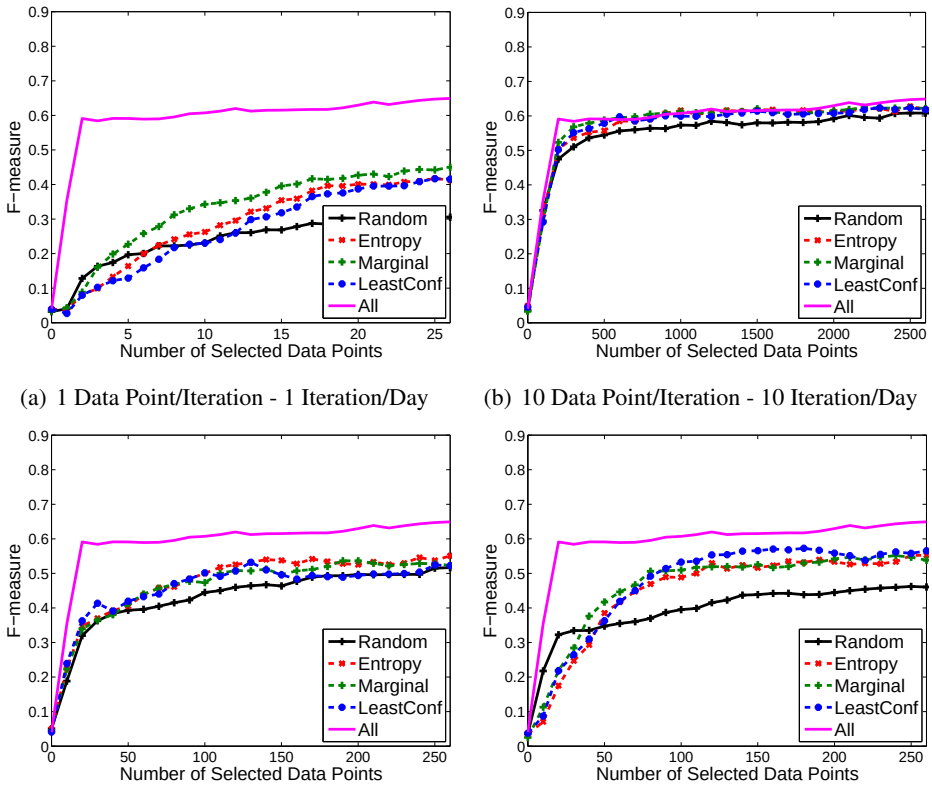
(d) 10 Data Point/Iteration - 1 Iteration/Day

Figure 6: Active learning experiment results for House B - Resident 1.

summary, even with the most demanding active annotation setup, the residents are queried 30% - 75% less and they are asked only a few number of data points in order to achieve the same recognition performance.

4.4 Discussion

It has been shown that active learning works well for an activity recognition application with experiments on real world data sets. With the active learning framework, the activity recognition system selects the most informative points. Then, the system is trained iteratively, using only the most informative points' labels. In the experiments, the points that needed to be annotated are selected on a daily basis. At the end of each day, the system asks the user what he/she has been doing during the time slices that are chosen to be the most informative. In the proposed scenario, it is possible that the user is disturbed only once a day, possibly before going to bed, by the system and asked about some activities he/she performed during that day. It is also possible that each iteration takes place at different times. This is important especially for the higher number of selections such as ten points in ten iterations



(a) 1 Data Point/Iteration - 1 Iteration/Day

(b) 10 Data Point/Iteration - 10 Iteration/Day

(c) 1 Data Point/Iteration - 10 Iteration/Day

(d) 10 Data Point/Iteration - 1 Iteration/Day

Figure 7: Active learning experiment results for House B - Resident 2.

cases. It could be difficult to obtain all 100 point in a single session.

The proposed active learning framework allows different number of data points to be selected from each day. Having more data points is always better but the number can vary from one to up to all data points. The model parameters are recalculated after each obtained label since each labeled point is of significant importance to obtain accurate model parameters. Since a supervised approach is used, recalculating the parameters is very fast and the user does not have to wait to be asked about the following label. The algorithm iteratively selects points and updates the model parameters, therefore, the bias on the selection do not propagate. Also, since always the true labels are obtained for the selected points, the bias on learning the model parameters is very unlikely to occur.

4.4.1 Random vs. Uncertainty Sampling

In nearly all of the cases, random selection performs worse than active learning methods. The exceptions occur especially with an extremely low number of data points. When the number of data points are too low, the model is not accurate

enough to correctly determine the importance of the data points. In that case, it is possible to come up with a higher performance with a random selection. Even with a random selection, the labels obtained are the ground truth labels so that they are useful in learning as well. However, in all the experiments, there is a clear distinction with the uncertainty measure based selection and random selection stating that these measures work better than random.

4.4.2 Comparison among Uncertainty Measures

As long as the different measures are concerned, there are not any significant differences. Nevertheless, marginal selection method has a slightly higher performance than the others whereas entropy method has a slightly lower performance than the others. In order to have a better understanding of the differences in the three measures, consider the following example probability distributions given in Table 1. In this scenario, three hypothetical time steps are considered for an activity recognition problem with four activity classes. The probability distributions provided by the current state of the algorithm are given in the table. Using these probability distributions, three different uncertainty measure scores for these time steps are calculated. In this scenario, if the entropy measure was used, the first time step would be selected for annotation. Likewise, if the least confident measure was used, t_2 would be selected. If the measure was margin sampling, then t_3 would be selected for annotation. In the table for probability distributions, the different behaviors of these measures as well as similarities can be observed. The margin sampling method considers only the two most likely classes and selects the data points that the difference between these two classes are minimum. The least confident measure, on the other hand, judges the uncertainty based on the most likely class only. When the model gives lower probabilities for the most likely class for some data points, then these data points are more likely to be selected by the least confident method. the entropy measure is the only measure in which all of the probabilities are taken into account. The entropy measure selects the data points for which the probability distribution is more uniform than the others. The entropy measure is maximized if all the probabilities for different classes are equal for a given data point.

Also, it is important to note that the measures do not select mutually exclusive data points. There are overlaps between the regions of probability distributions selected by these measures. For example, with a uniform distribution for a given data point (i.e. the case in which all the probabilities in the example are 0.25), all three measures will select this same data point for annotation since it is the most informative one for every measure.

In the experiments, it has been shown that entropy measure does not perform as well as the other methods. The probable cause of this lies in the nature of activity recognition problem. Since the entropy measure tends to select the instances in which every activity is nearly equally possible it tends to select instances from the *IDLE* class. The *IDLE* class is composed of time slices that are not labeled as any

Table 1: Example probability distributions and calculated scores

	t_1	t_2	t_3
$p(A_1)$	0.29	0.28	0.49
$p(A_2)$	0.24	0.27	0.49
$p(A_3)$	0.24	0.24	0.01
$p(A_4)$	0.23	0.21	0.01
Calculated Scores			
Entropy	**0.60**	0.59	0.34
LeastConf	0.71	**0.72**	0.51
Margin	0.05	0.01	**0.00**

other activity. This class acts like a transition state between two activities. For example, there can be *IDLE* time slices between *Sleeping* activity and *Brushing teeth* activity. The *IDLE* time slices constitutes the walking duration between the bedroom and bathroom in this case. Many of the data points in this class will look alike to other activities yielding a more uniform probability distribution in classification. Since these time slices are selected for annotation when the entropy metric is used, it is likely that the class labels obtained will belong to this single class. On the other hand, the classification performance can only be improved only if the model learns a more diverse set of activities. When margin sampling is used, on the other hand, the chance of selecting different activities increases. For instance, assume that the model asks for the label of a data point for which it assigned high and nearly equal probabilities for two similar activities such as *Brushing teeth* and *Shaving*. Once the label is obtained, the model has the chance of updating its parameters for both of these activities and not just one of them. If the actual label is *Shaving* for instance, the model has a better understanding of what *Shaving* activity looks like and also it knows more on what *Brushing teeth* activity does not look like. In another words, the model learns not to confuse between two similar activities with marginal sampling method and this has an boosting effect on the overall performance of the learning algorithm

4.4.3 Single iteration vs. Multiple iterations

In the results, two different configurations for ten points per day selection are provided, i.e. collecting ten points in a single iteration as opposed to a single point in ten iterations. Similarly, experiments that select 100 points per day in a single iteration and selecting ten points in ten iterations are conducted. Since the results for the former configuration are not so different from its ten iteration counterpart, the performance graphs of these experiments are not provided separately. The general performance trends are the same for both configurations in each case. On the other hand, when more iterations are made, the learning curve is steeper especially for the first few iterations. As the number of labeled points increase, the effect of iterations disappear. A steeper learning curve is expected since before asking for

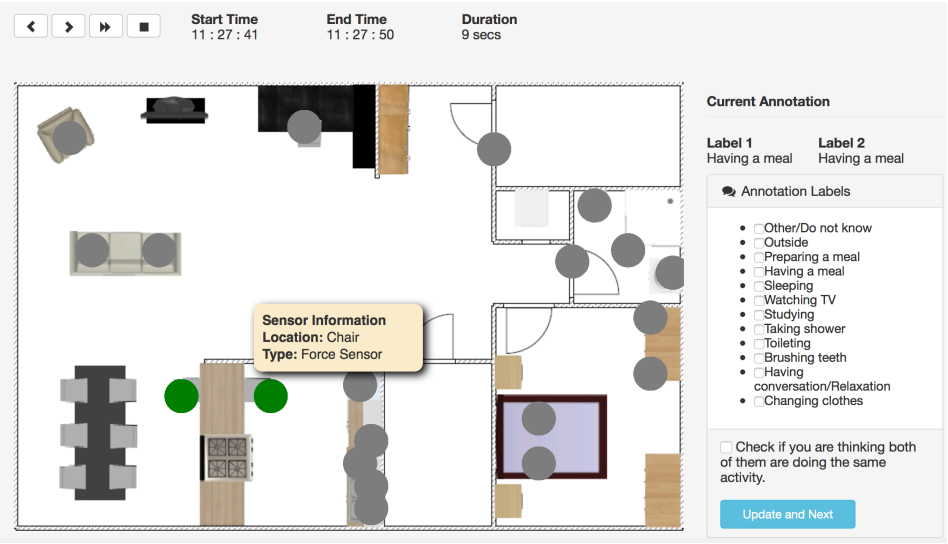


Figure 8: A screen shot from the web-based annotation tool.

new labeled points, the algorithm has the opportunity to update the model and ask about more informative labels with a more mature model. When a better model is obtained, the marginal benefit of the uncertainty measures increases. This, in turn, leads to a steeper increase in the performance in the beginning of active learning.

The number of iterations becomes an important design consideration if the learning algorithm has a high time complexity. If the running time of the learning algorithm is long enough to be noticed by the annotator then there will be pauses between the consecutive queries. If the pauses are too long then the annotator will become annoyed. In these cases, the iteration counts should be kept at minimum for usability purposes. When the learning algorithm is fast enough, using more iterations is more beneficial for efficient learning of the activities actively. Also, a hybrid approach could be employed since the effect of the higher number of iterations are shown to be more significant in the beginning.

5 Web-based Annotation Tool

One important concern with offline active annotation is the memory limitations of human annotators. Since the active query selection is performed after the whole day is over in the proposed scenario, a prototype application has been developed to mitigate the negative effects of incorrect retrieval. The proposed application can be used for both querying the annotator and also visualizing the sensor data for that specific moment.

In Figure 8, a sample screen is given from the developed sensor data annotation and visualization tool. This web-based simple yet efficient tool allows us to collect the necessary ground truth labels while also serving as a memory aid tool.

The application shows active and passive sensors together with their locations and types. While active sensors are shown as green circles, passive sensors are shown as grey circles. When the annotator moves the mouse over the circles, she/he can see location and type of sensors. This property helps annotators make better interpretation. After the users see this visualization of active and passive sensors, they are expected to annotate the activities choosing labels from set of activity labels on the right hand side. In order to further facilitate the retrieval process, the start and end times of the specified sensor state configuration together with the sensor firing duration information is provided at the top portion of the screen. Besides, the navigation buttons help the annotator to move back and forth between the time steps. This mechanism helps in making the temporal connections between consecutive time steps and in case of sensor failures or noisy firings, the annotator can make better interpretations about the ground truth activity labels.

One important benefit of having a web-based tool for annotation is that it allows utilizing other people for the annotation task. This feature may become useful especially in cases where the residents are incapable of annotating their own activities due to dementia or other diseases. In that case, authorized relatives or health care personnel can perform the annotation tasks. Although the accuracy is expected to be lower when compared to self annotation, the preliminary experiments with several unfamiliar annotators indicate a relatively high accuracy values for most activities of daily living such as sleeping, having a meal, toileting and watching TV. Activities that are more open to different interpretations such as relaxing or working are more challenging for unfamiliar annotators. Nevertheless, the flexibility of the overall learning phenomena makes it a proper candidate for large scale deployments of activity recognition systems.

The performance of web-based annotation tool is evaluated in [34]. According to the results from ten different people who have not participated in the data collection phase, the average recognition rate is found to be 70% and 64.5% for House A and House B, respectively. When the experiment is repeated with the actual residents who lived in the houses, the accuracies are as high as 95%. The main sources of errors were due to noisy sensor firings and the lack of temporal connections between the activities and also differences in the ways the activities performed. For example, the recognition rates for activities that are performed in specific locations in the house such as sleeping, having a shower, preparing a meal were considerably higher than other activities that are more variable in terms of locations. Due to the lack of a common pattern for watching TV or reading a book activities, it is harder to obtain the correct annotations from complete strangers. On the other hand, the residents themselves have higher annotation accuracy levels even for these leisure activities, making the annotation tool a viable solution for people who are familiar with the environment and behavior of the residents. Also, further improvements are possible with a tutoring mode for complete strangers by guiding them on how watching TV activity looks like.

6 Conclusion

In this paper, the scalability problems of automated human activity recognition systems are addressed since they require labeled data sets for adapting themselves to different users and environments. Collecting the data, annotation and retraining the systems from scratch for every person or every house is too costly. Therefore, redeploying these systems in different settings should be accomplished in a cost effective and user friendly way. For this purpose, active learning methods which reduce the annotation effort by selecting only the most informative data points to be annotated are proposed. In the proposed framework, user friendliness is also considered. It has been shown that by disturbing the user only a few times each day for obtaining the minimum amount of labels, it is possible to learn accurate model parameters.

Three different measures of uncertainty for selecting the most informative data points are used and their performance are evaluated by using real world data sets. Experiments showed that all three proposed method works well for the activity recognition system rather than random selection. By using the active learning, the annotation effort is reduced by a factor of two to four, depending on the house and resident setting in ARAS data sets.

Achieving high performance in activity recognition systems using probabilistic models depends on model parameters that are learned using the labeled data. With active learning, the aim is reaching the most accurate model parameters iteratively using the parameters obtained from previous iterations for selecting the most informative data points. In the first iterations, the parameters are based on few number of data points, therefore, not accurately estimated. This leads to a poor estimate of the informativeness of data points at the first iterations. The results suggest that even with a small amount of training data obtained after a few iterations, the selection gets better quickly. Therefore, instead of randomly initializing the parameters in the first iteration, transfer learning which allows the use of model parameters that have been learned previously to be used in another setting [35] can be used. As a future study, using transfer learning together with active learning methods could be explored to lead better estimates of the parameters even at the first iterations.

Acknowledgments

This work is supported by the Boğaziçi University Research Fund under the grant number BAP 8684.

References

[1] Hande Alemdar and Cem Ersoy. Wireless Sensor Networks for Healthcare: A Survey. *Computer Networks*, 54(15):2688–2710, 2010.

[2] Athanasios Bamis, Dimitrios Lymberopoulos, Thiago Teixeira, and Andreas Savvides. The BehaviorScope Framework for Enabling Ambient Assisted Living. *Personal Ubiquitous Computing*, 14(6):473–487, 2010.

[3] Albert Ali Salah, Theo Gevers, Nicu Sebe, and Alessandro Vinciarelli. Challenges of Human Behavior Understanding. In *First International Conference on Human Behavior Understanding*, HBU '10, pages 1–12, 2010.

[4] Michael C. Mozer. The Neural Network House: An Environment that Adapts to its Inhabitants. In *AAAI Spring Symposium on Intelligent Environments*, pages 110–114, 1998.

[5] Sumi Helal, William Mann, Hicham El-Zabadani, Jeffrey King, Youssef Kaddoura, and Erwin Jansen. The Gator Tech Smart House: A Programmable Pervasive Space. *Computer*, 38(3):50–60, 2005.

[6] Gregory D Abowd, Aaron F Bobick, Irfan A Essa, Elizabeth D Mynatt, and Wendy A Rogers. The Aware Home: A Living Laboratory for Technologies for Successful Aging. In *AAAI-02 Workshop Automation as Caregiver*, 2002.

[7] Minoh Michihiko and Tatsuya Yamazaki. Daily life support experiment at ubiquitous computing home. In *11th Information Processing and Management of Uncertainty in Knowledge-Based Systems International Conference*, 2006.

[8] Taketoshi Mori, Aritoki Takada, Hiroshi Noguchi, Tatsuya Harada, and Tomomasa Sato. Behavior prediction based on daily-life record database in distributed sensing space. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1703–1709. IEEE, 2005.

[9] Tatsuya Yamazaki. Ubiquitous home: real-life testbed for home context-aware service. In *First International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, 2005. Tridentcom 2005.*, pages 54–59. IEEE, 2005.

[10] Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson. Activity Recognition in the Home Using Simple and Ubiquitous Sensors. In *International Conference on Pervasive Computing*, Pervasive '04, pages 158–175, 2004.

[11] Geetika Singla, Diane J. Cook, and Maureen Schmitter-Edgecombe. Recognizing Independent and Joint Activities Among Multiple Residents in Smart Environments. *Journal of Ambient Intelligence and Humanized Computing*, 1(1):57–63, 2010.

[12] Tim van Kasteren. *Activity Recognition for Health Monitoring Elderly Using Temporal Probabilistic Models*. PhD thesis, University of Amsterdam, Netherlands, 2011.

[13] Fco. Javier Ordonez, Paula de Toledo, and Araceli Sanchis. Activity Recognition Using Hybrid Generative/Discriminative Models on Home Environments Using Binary Sensors. *Sensors*, 13(5):5460–5477, 2013.

[14] Iram Fatima, Muhammad Fahim, Young-Koo Lee, and Sungyoung Lee. A Unified Framework for Activity Recognition-Based Behavior Analysis and Action Prediction in Smart Homes. *Sensors*, 13(2):2682–2699, 2013.

[15] Mukhtiar Memon, Stefan Rahr Wagner, Christian Fischer Pedersen, Femina Hassan Aysha Beevi, and Finn Overgaard Hansen. Ambient Assisted Living Healthcare Frameworks, Platforms, Standards, and Quality Attributes. *Sensors*, 14(3):4312–4341, 2014.

[16] B. Settles and M. Craven. An Analysis of Active Learning Strategies for Sequence Labeling Tasks. In *Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, 2008.

[17] B. Anderson, S. Siddiqi, and A. Moore. Sequence Selection for Active Learning. Technical report, Carnegie Mellon University, 2006.

[18] Shusen Zhou, Qingcai Chen, and Xiaolong Wang. Active deep networks for semi-supervised sentiment classification. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 1515–1523. Association for Computational Linguistics, 2010.

[19] Dan Wang and Yi Shang. A new active labeling method for deep learning. In *International Joint Conference on Neural Networks (IJCNN)*, pages 112–119. IEEE, 2014.

[20] Heng-Tze Cheng, Feng-Tso Sun, Martin Griss, Paul Davis, Jianguo Li, and Di You. Nuactiv: Recognizing unseen new activities using semantic attribute-based learning. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 361–374. ACM, 2013.

[21] T. Truyen, H. Bui, D. Phung, and S. Venkatesh. Learning Discriminative Sequence Models from Partially Labelled Data for Activity Recognition. In *PRICAI 2008: Trends in Artificial Intelligence*, pages 903–912, 2008.

[22] M. Hasan and A.K. Roy-Chowdhury. A continuous learning framework for activity recognition using deep hybrid feature models. *IEEE Transactions on Multimedia*, PP(99):1–1, 2015.

[23] B. Settles. *Active Learning*. Morgan&Claypool, 2012.

[24] R. Liu, T. Chen, and L. Huang. Research on Human Activity Recognition Based on Active Learning. In *International Conference on Machine Learning and Cybernetics*, ICMLC '10, pages 285–290, 2010.

[25] M. Stikic, K. van Laerhoven, and B. Schiele. Exploring Semi-supervised and Active Learning for Activity Recognition. In *12th IEEE International Symposium on Wearable Computers*, ISWC '08, pages 81–88, 2008.

[26] Y. Ho, C. Lu, I. Chen, S. Huang, C. Wang, and L. Fu. Active-learning Assisted Self-reconfigurable Activity Recognition in a Dynamic Environment. In *IEEE International Conference on Robotics and Automation*, pages 813–818, 2009.

[27] Liyue Zhao, Gita Sukthankar, and Rahul Sukthankar. Robust active learning using crowdsourced annotations for activity recognition. In *Human Computation*, 2011.

[28] Walter S Lasecki, Young Chol Song, Henry Kautz, and Jeffrey P Bigham. Real-time crowd labeling for deployable activity recognition. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 1203–1212. ACM, 2013.

[29] E. Hoque and J. Stankovic. AALO: Activity Recognition in Smart Homes Using Active Learning in the Presence of Overlapped Activities. In *6th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '12, pages 139 –146, 2012.

[30] Salikh Bagaveyev and Diane J Cook. Designing and evaluating active learning methods for activity recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 469–478. ACM, 2014.

[31] D.D. Lewis and J. Catlett. Heterogeneous Uncertainty Sampling for Supervised Learning. In *11th International Conference on Machine Learning*, ICML '94, pages 148–156, 1994.

[32] L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[33] Hande Alemdar, Halil Ertan, Ozlem Durmaz Incel, and Cem Ersoy. ARAS Human Activity Datasets in Multiple Homes with Multiple Residents. In *7th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '13, pages 232–235, 2013.

[34] Nezihe Pehlivan, Hande Alemdar, Can Tunca, and Cem Ersoy. Human Activity Recognition and Interpretation in Smart Home: An Annotation and Data Visualization Tool. In *Akademik Bilişim*, AB '15, Eskişehir, Turkey, 2015.

[35] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse. Transferring Knowledge of Activity Recognition Across Sensor Networks. *Pervasive Computing*, pages 283–300, 2010.