Mindful Active Learning

Zhila Esna Ashari[*] and **Hassan Ghasemzadeh**

School of Electrical Engineering & Computer Science, Washington State University, Pullman, WA
{z.esnaashariesfahan, hassan.ghasemzadeh}@wsu.edu

Abstract

We propose a novel active learning framework for activity recognition using wearable sensors. Our work is unique in that it takes physical and cognitive limitations of the oracle into account when selecting sensor data to be annotated by the oracle. Our approach is inspired by human-beings' limited capacity to respond to external stimulus such as responding to a prompt on their mobile devices. This capacity constraint is manifested not only in the number of queries that a person can respond to in a given time-frame but also in the lag between the time that a query is made and when it is responded to. We introduce the notion of *mindful active learning* and propose a computational framework, called *EMMA*[1], to maximize the active learning performance taking informativeness of sensor data, query budget, and human memory into account. We formulate this optimization problem, propose an approach to model memory retention, discuss complexity of the problem, and propose a greedy heuristic to solve the problem. We demonstrate the effectiveness of our approach on three publicly available datasets and by simulating oracles with various memory strengths. We show that the activity recognition accuracy ranges from 21% to 97% depending on memory strength, query budget, and difficulty of the machine learning task. Our results also indicate that EMMA achieves an accuracy level that is, on average, 13.5% higher than the case when only informativeness of the sensor data is considered for active learning. Additionally, we show that the performance of our approach is at most 20% less than experimental upper-bound and up to 80% higher than experimental lower-bound. We observe that mindful active learning is most beneficial when query budget is small and/or oracle's memory is weak, thus emphasizing contributions of our work in human-centered mobile health settings and for elderly with cognitive impairments.

1 Introduction

Coupled with machine learning algorithms, mobile and wearable devices are being increasingly utilized to monitor physical and mental health, and to improve human wellbeing through clinical interventions. To train accurate machine learning models, an adequate number of labeled sensor data is required. Unfortunately, models that are trained based on sensor data collected in controlled environments and laboratory settings perform extremely poorly when utilized in uncontrolled environments and outside clinics [Fallahzadeh and Ghasemzadeh, 2017].

For the machine learning algorithms to be accurate, one needs to collect and label sensor data in end-user settings. Therefore, active learning is a natural choice for labeling the data where the end-user[2] acts as the oracle agent and we iteratively query the user for correct labels [Settles, 2012; Bachman *et al.*, 2017]. In such a human-centered setting, it is critical to design active learning strategies that are mindful of the user's physical and cognitive capabilities. We recognize that human beings are limited in their capacity to respond, for example, to prompts on their mobile devices. This capacity constraint is usually manifested in the number of queries that a person can respond to in a given time-frame and in the difference between the time that a query is made and when it has been responded to. This issue is critical in wearable-based continuous health monitoring where the amount of sampled data is orders of magnitude more than what the end-user can possibly annotate [Redmond *et al.*, 2014; Ma *et al.*, 2019]. In this paper, we introduce the notion of *mindful active learning* and propose *EMMA* to maximize the active learning performance taking informativeness of sensor data, query budget, and human memory into account. To the best of our knowledge, our work is the first study that combines informativeness of queried data with oracle's memory strength in a unified framework for active learning.

1.1 Related Work

Active learning has been widely used for human-in-loop learning tasks to iteratively interact with an expert user to retrieve necessary information for improved learning performance [Huang *et al.*, 2014; Sabato and Hess, 2016;

---

[*]Contact Author

[2]The terms 'end-user' and 'oracle' are interchangeably used in this paper.

Xu *et al.*, 2007; Jun and Nowak, 2016; Hoi *et al.*, 2006; Shen and Zhai, 2005; Sener and Savarese, 2018]. Prior research studied the problem of active learning under constrained query budget for image and video recognition [Vijayanarasimhan *et al.*, 2016]. Active learning has been also studied in the context of activity recognition [Stikic and Schiele, 2009; Diethe *et al.*, 2016]. For instance, accelerometer data annotated by end-users are used to detect physical activities [Bao and Intille, 2004]. Active learning has been used for many other human-centered prediction problems beyond those that use wearable sensors and images [Settles, 2012; Murukannaiah and Singh, 2015]. However, prior research does not take into account important cognitive attributes of human beings such as memory strength and forgetfulness of the events while designing active learning solutions. Current research makes an implicit assumption that either the oracle has a perfect memory that can precisely remember all the events, or each query is instantaneously responded to. None of these assumptions is realistic in continuous health monitoring settings where the end-user also acts as our oracle. In this paper, we attempt to take the first steps at integrating physical and cognitive attributes of the oracle with active learning. In particular, we account for (i) the oracle's physical capability to respond to active learning queries as measured by the number of queries that are made; and (ii) the oracle's cognitive capacity to respond to the queries as measured by memory retention, a function that combines memory strength of the oracle agent with the amount of delay in responding to the issued queries.

1.2 Contributions

Our contributions in this paper can be summarized as follows: (i) we introduce mindful active learning as a general approach for budget-aware and delay-tolerant active learning in human-in-the-loop health monitoring applications. Mindful active learning takes into account the possibility of forgetting the event of interest by the oracle based on the time difference between querying and the activity being performed constrained on a maximum number of queries; (ii) we propose an approach to model memory retention based on the Ebbinghaus forgetting curve [Ebbinghaus, 1885]; (iii) we formulate mindful active learning as an optimization problem and show that the problem is NP-hard; (iv) we propose a greedy heuristic algorithm to solve the optimization problem; and (v) we evaluate the performance of EMMA on three datasets for activity recognition where the goal is to detect physical activities performed by the end-user.

2 Problem Statement

Assume that we are given a collection of sensor measurements recorded while wearable sensors are carried by the user during daily activities. Without loss of generality, we assume that the sensor measurements, referred to as sensor observations henceforth, are represented in a feature space. These observations need to be used to train a classifier for activity recognition. Because the sensor observations are unlabeled, we use active learning to construct a labeled training set by querying the end-user to annotate a subset of the observations.

Because of the lag between querying a sensor observation and performing the activity associated with that observation, it is possible that the oracle is incapable of remembering the correct label. We assume that the likelihood of such a mislabeling is a function of the oracle's memory strength and the time lag. Moreover, we assume that the number of allowed queries is constrained to a given value, referred to as *budget*, in order to minimize burden of the oracle. Therefore, we need to select a subset of sensor observations, upper-bounded by the given budget, such that the probability of obtaining correct and informative labels is maximized. This problem can be defined as follows.

Problem 1 (Mindful Active Learning) *Let $\mathcal{A} = \{A_1, A_2,..., A_n\}$ represent the set of activities that need to be recognized by the wearable system. We refer to this set as activity vocabulary. Furthermore, let $\mathcal{X} = \{X_1, X_2,..., X_m\}$ be a set of m observations made by the sensors at times $\mathcal{T} = \{t_1, t_2,..., t_m\}$. The task of active learning is to query the user, with a predefined memory strength of s, at time $t_q \geq t_m$, to label sensor observations in $\mathcal{X}$ and train an activity recognition model using the labeled observations. The active learning process is constrained by limiting the number of queries to a given upper-bound budget, B.*

2.1 Problem Formulation

The task of selecting a subset of observations to be labeled by the oracle can be viewed as finding a set $\mathcal{Z} = \{z_1, z_2,..., z_k\}$ of k observations in $\mathcal{X}$ (i.e., $\mathcal{Z} \subseteq \mathcal{X}$ and $k \leq B$) such that the misclassification error due to a model trained over $\mathcal{Z}$ is minimized. The goal is to select the best subset of unlabeled observations to form a candidate query set (Z) that results in an accurate classifier. In order to maximize the performance of the final classifier, we consider two criteria including informativeness of the candidate observation and the oracle's ability to remember the correct label at time t_q. We use I_i and M_i to refer to informativeness and memory measures for a given observation X_i captured at time t_i.

This optimization problem can be formulated as

$$\text{Maximize} \quad \sum_{i=1}^{m} a_i \mathcal{E}(I_i, M_i) \quad (1)$$

Subject to:

$$X_i \in \mathcal{X} \quad (2)$$

$$\sum_i a_i \leq B \quad (3)$$

$$a_i \in \{0, 1\} \quad (4)$$

where X_i denotes the senor observation captured at time t_i, and B represents the budget. The binary variable a_i determines whether or not observation X_i is selected for inclusion in $\mathcal{Z}$. The objective function in (1) attempts to maximize the total amount of expected gain ($\mathcal{E}$) given informativeness and memory for individual observations.

We propose to use uncertainty of the model with respect to a given observation as a measure of informativeness of that

observation for inclusion in $\mathcal{Z}$. We also propose to measure memory by combining memory strength of the user with the expected error due to difference in time between query issuance and activity occurrence. The expected gain is then computed as

$$\mathcal{E}(I_i, M_i) = \mathcal{E}(I_i)\mathcal{E}(M_i) \tag{5}$$

To quantify informativeness (I_i) of observation X_i, we propose to use entropy, as shown in (6), to measure how certain the model is about its predicted label for X_i.

$$\mathcal{E}(I_i) = E_i = -\sum_{j=1}^{n} P_{ij} \log P_{ij} \tag{6}$$

The term E_i in (6) refers to entropy for observation X_i, and P_{ij} represents the probability of X_i being classified as A_j. Because the classifier is less certain to classify observations that carry a higher entropy, such observations will naturally be more informative if labeled and used for classifier retraining. Therefore, E_i is a reasonable proxy for $\mathcal{E}(I_i)$. To quantify memory, M_i, for observation X_i, we define memory retention as follows.

Definition 1 (Memory Retention) *Memory retention, R, is defined as the probability of a human subject with a memory strength of s being able to remember an event correctly after a given time, t, has elapsed.*

We use Ebbinghaus forgetting curve [Ebbinghaus, 1885] to quantify memory retention. To this end, memory retention for observation X_i is given by

$$\mathcal{E}(M_i) = R_i = e^{-\Delta t_i/s} \tag{7}$$

where Δt_i denotes the difference in time between occurrence of the event represented by observation X_i and issuance of the query (i.e., t_q). Furthermore, the term s represents memory strength, which is a user-specific value. The memory retention in (7) is naturally a measure for receiving correct labels from the oracle. Therefore, the memory retention R_i can be used to quantify the memory for observation X_i (i.e., $\mathcal{E}(M_i) = R_i$).

Problem 2 (Entropy-Memory Maximization) *By replacing $\mathcal{E}(I_i)$ and $\mathcal{E}(M_i)$ in (5) with entropy measure in (6) and memory retention in (7), we can rewrite the objective function in (1) as follows.*

$$\text{Maximize} \quad \sum_{i=1}^{m}\sum_{j=1}^{n} a_i e^{-\Delta t_i/s}(-P_{ij} \log P_{ij}) \tag{8}$$

We note that this new objective function in (8) is subject to the same constraint in (2)–(4). We refer to this formulation of mindful active learning as Entropy-Memory Maximization (EMMA).

2.2 Problem Complexity

In this section, we discuss complexity of the entropy-memory maximization. We first show that the problem is NP-hard.

Theorem 1 *The entropy-memory maximization in Problem 2 is NP-Hard.*

Proof 1 *Proof by reduction from the well-known Knapsack problem [Pisinger, 2005]. Details are eliminated for brevity.*

To provide some insight into the input size of the EMMA problem, consider a naive approach (i.e., brute-force), in which we cycle through all subsets of the unlabeled dataset with the number of elements less than or equal to B. For each subset, we can generate all possible orderings of the sensor observations within that subset. For each ordering, we draw one observation from the subset at a time, compute entropy and memory retention values, query the oracle, and train a classifier with the observations labeled so far. By repeating this process for all subsets and all orderings of the observations within each subset, we can finally choose the ordering and subset that achieve the maximum value for the objective function in (8). Although this brute-force approach is impractical for real-world deployment, an analysis of the complexity of such an approach provides insight into the relationship between the complexity function and various problem parameters.

Theorem 2 *The time complexity of the brute-force approach for EMMA is exponential in the query budget, B.*

Proof 2 *It is straightforward to see that the overall time complexity of the brute-force solution is $O(|S|)$ where $|S|$ refers to the total number of orderings of all subsets of unlabeled set $\mathcal{X}$ with a subset size less than or equal to B. For each subset of size b, there exist $b!$ orderings of the observations that reside within the subset. Therefore, $|S|$ is given by*

$$|S| = \sum_{b=1}^{B}\binom{m}{b} \times b! = \sum_{b=1}^{B}\frac{m!}{(m-b)!} \tag{9}$$

where m denotes the size of the unlabeled set $\mathcal{X}$. The equation in (9) can be presented as

$$|S| = \sum_{b=1}^{B}\prod_{k=1}^{b}(m-b+k) \tag{10}$$

Because in real-world scenarios, the number of recorded sensor observations is orders of magnitude higher than the number of queries that the oracle can respond to (i.e., $B \ll m$), we can write

$$\prod_{k=1}^{b}(m-b+k) \approx m^b \tag{11}$$

Therefore, (10) can be rewritten as

$$|S| \approx m + m^2 + ... + m^B \approx \frac{m^{(B+1)} - 1}{m-1} \approx m^B \tag{12}$$

Therefore, the time complexity of the naive approach is $O(m^B)$.

Algorithm 1 Greedy algorithm for EMMA

Input: $\mathcal{X}$ (unlabeled observations), B (budget)
Output: $\mathcal{Z}$ (labeled observations)
for $b = 1$ to B **do**
    Compute $\mathcal{E}(I_i, M_i)$ for all $X_i \in \mathcal{X}$ using (5)–(7)
    Find $X_i \in \mathcal{X}$ with highest value of $\mathcal{E}(I_i, M_i)$
    Remove X_i from $\mathcal{X}$
    Query oracle to annotate X_i, and add labeled X_i to $\mathcal{Z}$
    Retrain model $\mathcal{M}$ using labeled items in $\mathcal{Z}$
end for

---

3 Problem Solution

Recall that the goal of the optimization problem in EMMA is to select a subset of observations of high-entropy such that their labels are likely of being remembered by the oracle. Because the problem is NP-hard in nature, we design a greedy heuristic algorithm, shown in Algorithm 1, to solve the problem.

This simple greedy algorithm iteratively chooses the best candidate observation from the set of unlabeled observations $\mathcal{X}$ with the highest expected gain, shown in (5), assuming that the informativeness and memory are measured as given in (7) and (8). Note that after moving an observation from $\mathcal{X}$ to $\mathcal{Z}$, model $\mathcal{M}$ is retrained using the labeled data in $\mathcal{Z}$. This procedure is repeated until the entire budget is consumed.

Theorem 3 *The greedy approach shown in Algorithm 1 provides a $\frac{1}{2}$-approximation for EMMA.*

Proof 3 *Proof is eliminated for brevity.*

Theorem 4 *The time complexity of Algorithm 1 is linear in m, the number of original unlabeled observations in $\mathcal{X}$.*

Proof 4 *The 'for' loop iterates for 'B' times. During each iteration, we need to (i) compute $\mathcal{E}(I_i, M_i)$ for the remaining elements in $\mathcal{X}$; and (ii) find $X_i \in \mathcal{X}$ with highest value of $\mathcal{E}(I_i, M_i)$. Both of these operations require $O(|\mathcal{X}|)$ to complete assuming a constant time complexity for computing $\mathcal{E}(I_i, M_i)$ for a given X_i. Therefore, the time complexity of Algorithm 1 is $B \times O(|\mathcal{X}|)$. Because $|\mathcal{X}|$ is initially m and decreases by one in each iteration, the total number of times to compute $\mathcal{E}(I_i, M_i)$ is given by*

$$\sum_{b=1}^{B}(m - b - 1) = O(m.B) \tag{13}$$

Therefore, the time complexity of Algorithm 1 is $O(m.B)$

4 Validation

This section presents our validation approach, results on the performance of our mindful active learning algorithms, and a comparative analysis. Throughout this section, we use *EMMA* to refer to the greedy algorithm presented in Section 3. We note that the naive approach discussed in Section 2.2 has an exponential time complexity, which makes it impractical given the large amounts of data collected during continuous health monitoring using wearable sensors.

4.1 Experimental Setup

To assess the performance of EMMA, we used three real-world, publicly-available, sensor-based datasets including HART [Anguita *et al.*, 2013], DAS [Altun *et al.*, 2010; Barshan and Yüksek, 2014; Altun and Barshan, 2010], and AReM [Palumboa *et al.*, 2016]. HART contained sensor data collected with 30 human subjects during six activities. For DAS, data from eight human subjects and 10 activity classes were used. The AReM dataset involved six activities and one human subject.

We also simulated oracles with various memory strengths. The goal was to measure the likelihood of an incorrect label provided by the oracle given the memory strength and the time that has passed from capturing the queried observation. To simulate the oracle's remembering of the event associated with a queried sensor observation X_i based on a given memory strength value s, we first computed memory retention R_i for X_i using (7). We then assigned the correct label with the probability R and incorrect label with the probability $(1-R)$. To alleviate the effect of randomness in our simulation, we repeated each experiment 30 times and report the average results.

For comparative analysis, we assessed the performance of EMMA against the following active learning approaches:

- **EMA** (Entropy Maximization): this approach solves the optimization problem based on informativeness of the queried observations only, while still considering oracle's memory deficiency effects in responding to the queries. In this case, $\mathcal{E}(I_i)=E_i$ and $\mathcal{E}(M_i) = 1$.

- **MMA** (Memory Maximization): this algorithm aims to maximize the objective function for memory retention only. This means, the algorithm tends to choose observations whose probability of remembering the label is higher than others. In this case, $\mathcal{E}(I_i)=1$ and $\mathcal{E}(M_i)=R_i$.

- **Upper-bound (UB)**: this refers to the case where no erroneous labels exist as a result of memory deficiency. That is, we assume that the oracle's memory is perfect, as a result of which the optimization problem aims to maximize for entropy only.

- **Lower-bound (LB)**: this refers to the case where the oracle's memory is low and the observations are chosen randomly. Therefore, the informativeness of the queried observation is not considered as a parameter.

Throughout our experiments, we used SVM (Support Vector Machine) as our underlying activity recognition classifier. This choice was made based on extensive experiments on the datasets, where we considered four types of machine learning algorithms including SVM with linear kernel, SVM with RBF (Radial Basis Function) kernel, logistic regression, and decision tree. These algorithms achieved 92%, 83%, 90%, 88% average accuracy, respectively, over all subjects. Note that the presented methodologies are independent of choice of the activity recognition classifier. We also assumed that a small number of labeled samples are initially available in $\mathcal{Z}$. In our experiments, $\mathcal{Z}$ initially had two randomly selected labeled samples. This allowed us to learn an initial model to calculate primary entropy values for active learning.
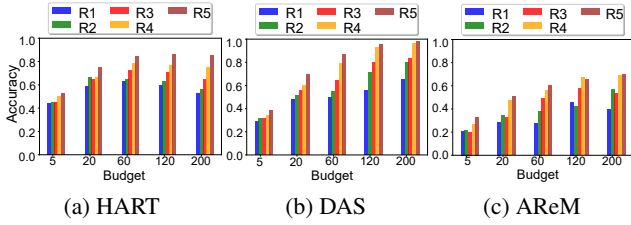
Figure 1: Accuracy of activity recognition model learned using EMMA on three datasets (HART, DAS, and AReM) as a function of query budget for five different memory retention levels (R_1–R_5).
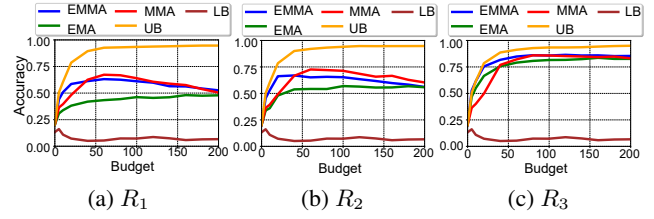

Figure 2: Activity recognition accuracy of various active learning algorithms on HART dataset using three different memory retention levels, R_1 (low), R_2 (medium), and R_3 (high).

4.2 Performance of EMMA

As a first analysis, we evaluated the performance of EMMA and examined how its accuracy changes as various parameters of the algorithm change. To this end, we conducted multiple experiments by changing algorithm parameters including query budget, B, and memory strength, s. After EMMA constructed a labeled dataset, we trained an activity recognition classifier using the labeled data.

Figure 1 shows the accuracy of the learned classifier on the three datasets as the query budget ranges from 5 to 200. The graphs show the accuracy values for five different memory retention levels including R_1 (10%–99%), R_2 (20%–99%), R_3 (30%–99%), R_4 (50%–99%), and R_5 (70%–99%). Recall that the memory retention R refers to the probability of remembering label for a given sensor observation in the dataset. However, memory retention is also a function of both memory strength and time different between query issuance and labeling (i.e., lag or delay). We also note that sensor observations have different time stamps and therefore represent different time delays. Therefore, each memory retention value (i.e., R_1 to R_5) in Figure 1 us represented as a range rather than a fixed value. Note that, as the time window size to capture each sensor observation from raw sensor readings varies across the datasets, we used different memory strength values to obtain the same memory retention intervals/ranges to provide comparable results.

As Figure 1 suggests, the activity recognition accuracy improves for each memory retention value as the query allowance (i.e., budget) increases. Also, as the memory strength grows, the accuracy improves for each budget value. The minimum accuracy, achieved with least budget and weakest memory is 44%, 29% and 21% for HART, DAS and AReM datasets, respectively. The accuracy reaches its maximum values of 85.3%, 97.5% and 70% with greatest budget and strongest memory for the three datasets. An interesting observation from Figure 1 is the performance decline in weak memory cases. It is generally known that increasing the number of labeled observations using active learning improves the classifier performance. However, we can see that adding a large number of labeled observations may reduce the accuracy if the memory retention is low due to a weak memory. An example of such cases is shown in Figure 1a for retention levels R_1, R_2, and R_3. In all these cases, the activity recognition performance improves up to a certain point (e.g., B=60) and it starts dropping after that point, mainly because more incorrect labels are added to the dataset resulting in a less accurate model. For instance, the accuracy for R_1, as

the weakest memory, starts from 44% for B=5, reaches up to 63% for B=60, and again drops to 52.5% for B=200. Another interesting observation is that the accuracy saturates after certain point with strong memories. This can be observed, for example, in Figure 1a for memory retention levels R_4 and R_5 where the accuracy plateaus after acquiring 60 labeled observations and stays around 77% and 85% for higher budget values. The reason is that, in this dataset, each added observation is very informative and the classifier learns faster and needs fewer observations to achieve its maximum capability.

The difference in accuracy among different datasets in Figure 1 can be explained by the quality of the data and the difficulty of the recognition task. For example, the DAS dataset (i.e., Figure 1b) contains highly discriminative features over different activities. As a result, adding a few number of correctly labeled observations has a significant impact on improving the classifier performance. For instances, when considering R_5, the change in accuracy when B grows form 5 to 20 is 23% for HART and 30% for DAS. On the contrary, AReM dataset (i.e., Figure 1c) contains less discriminative features and the window size over which the features are calculated is small, leading to needing a lot more labeled observations to learn new activities. Therefore, the growth in the performance of the classifier in Figure 1c is slow compared to other datasets. Specifically, the amount of improvement in accuracy is 17% for R_5 when B increases form 5 to 20 in AReM dataset, while this improvement is 23% and 30% for HART and DAS datasets, respectively.

4.3 Comparative Analysis

We compared the performance of EMMA with that of several active learning approaches including EMA, MMA, upper-bound (UB), and lower-bound (LB). See Section 4.1 for a description of these alternative approaches. For brevity, we focus on accuracy as our main performance measure here. Similar to our analysis in the previous section, we examined the performance of each algorithm while the budget value ranged from 0 to 200 and the memory strength was set such that different memory retention levels, R_1 (10%–99%), R_2 (25%–99%), and R_3 (70%–99%) were obtained. The results of this comparative analysis are shown in Figure 2, Figure 3, and Figure 4 for HART, DAS, and AReM datasets, respectively. The accuracy numbers in these figures represent average values computed over all users in each dataset.

The results show that the performance improvement due to using EMMA over EMA and MMA is most notable when the amount of budget is small and the memory is weak (i.e.,
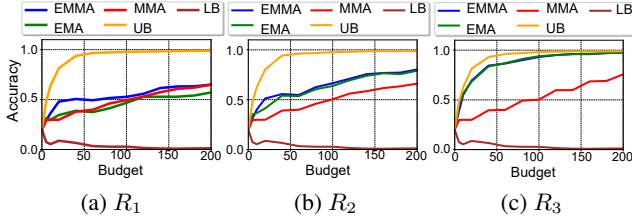
Figure 3: Activity recognition accuracy of various active learning algorithms on DAS dataset using three different memory retention levels, R_1, R_2, and R_3.

Figure 4: Activity recognition accuracy of various active learning algorithms on AReM dataset using three different memory retention levels, R_1, R_2, and R_3.

memory retention level of R_1). This observation emphasizes the importance of considering both uncertainty and memory retention in health monitoring applications where data collection is extremely costly and end-users may be cognitively impaired due to aging. It can be seen from the results that EMMA achieves an average accuracy that is 13.5% higher than that of EMA and 14% higher than that of MMA for cases of weaker memory and smaller budget. Moreover, we note that, for all datasets, as the memory becomes stronger, EMMA and EMA methods converge and achieve accuracy values closer to the upper-bound. Additionally, the performance of EMMA is at most 20% less than the experimental upper-bound and up to 80% higher than the experimental lower-bound, on average. Furthermore, for HART dataset, MMA converges with EMMA and EMA as well, while for other datasets, the performance of MMA improves very slower. The reason is that MMA tends to select the observations sequentially in time and ignores the informativeness of the observations. Therefore, if the oracle repeats an activity for a long time (e.g., 'sleeping', 'watching TV'), this method continues to querying for the same activity, resulting in a model that is incapable of recognizing other activities. Consequently, the active learning process needs to query for many new observations in order to learn new activities (see results for DAS and AReM datasets). Note, as mentioned previously, AReM contains less discriminative features compared to the two other datasets. This is the reason why the performance of MMA is closer to the lower-bound on AReM.

A notable observation is difference in the performances of MMA and EMMA on HART and AReM datasets as shown in Figure 2 and Figure 4. Specifically, MMA outperforms EMMA slightly in larger budgets in Figure 2, but it performs much worse than EMMA in Figure 4. This is because of switching fast between unknown activities in the HART dataset (Figure 2), which results in MMA learning new activities at a faster pace compared to how MMA learns in the AReM dataset (Figure 4). This observation suggests that the performance of MMA is highly dependent on the dataset (i.e., sequence of activities that occur).

We can also compare the performance of EMMA with that of EMA in these two datasets. As shown in Figure 2, EMMA outperforms EMA on HART. However, as shown in Figure 4, EMMA and EMA achieve a comparable performance on AReM. This can be explained as follows. Because activities in the HART dataset change frequently over time, considering time delays during optimization (i.e., in EMMA) contributes to an improved performance. In contrast, the per-
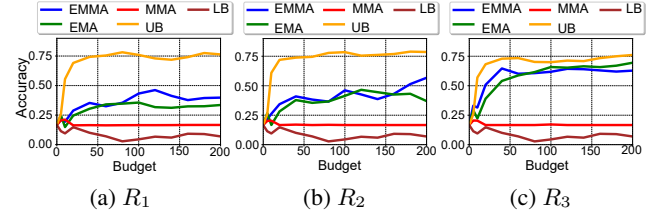
formance of EMA is highly dependent of the nature of the occurring events.

5 Conclusions and Future Work

Prior research on active learning takes informativeness of data and query budget into account when selecting the data for annotation. In this paper, we showed that cognitive constrains of the oracle are of significant importance that can greatly compromise the active learning performance. We posed an optimization problem to combine data uncertainty with memory retention for its use in ubiquitous and mobile computing applications. We showed that this problem is NP-hard and derived a greedy approximation algorithm to solve the proposed mindful active learning problem. Our extensive analyses on three publicly available datasets showed that EMMA achieves up to 97% accuracy for activity recognition using wearable sensors. We also showed that integrating memory retention improves the active learning performance by 16%.

Our results indicate that the performance of EMMA improves when oracle's memory is stronger or the query budget is higher. However, we noticed that increasing the budget does not improve the accuracy when highly accurate labels are available due to strong memory retention. We also noted that the active learning performance can decrease with increased budget if the oracle's memory is weak. Finally, the gap between the performance of EMMA and other algorithms is most notable with small budgets and weak memories.

In this paper, we focused on pool-based active learning. Our ongoing work involves developing mindful active learning strategies that make query decisions on-the-fly as wearable sensor data become available in real-time.

Our future work will also focus on conducting user studies that involve patients with cognitive impairments where we can assess the oracle's memory retention quantitatively. Moreover, in this paper, we assumed that the oracle responds to an issued query instantaneously. We plan to investigate active learning solutions that take into account the possibility of delayed responses through context-sensitive active learning.

Acknowledgments

References

[Altun and Barshan, 2010] Kerem Altun and Billur Barshan. Human activity recognition using inertial/magnetic sensor units. In *First International Workshop on Human Behavior Understanding*, Istanbul, Turkey, August 2010.

[Altun *et al.*, 2010] Kerem Altun, Billur Barshan, and Orkun Tunçel. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition*, 43(10):3605–3620, October 2010.

[Anguita *et al.*, 2013] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN*, Bruges, Belgium, April 2013.

[Bachman *et al.*, 2017] Philip Bachman, Alessandro Sordoni, and Adam Trischler. Learning algorithms for active learning. In *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, 2017.

[Bao and Intille, 2004] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, Linz/Vienna, Austria, April 2004.

[Barshan and Yüksek, 2014] Billur Barshan and Murat Cihan Yüksek. Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *The Computer Journal*, 57(11), 2014.

[Diethe *et al.*, 2016] Tom Diethe, Niall Twomey, and Peter Flach. Active transfer learning for activity recognition. In *Proceedings of proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, April 2016.

[Ebbinghaus, 1885] Hermann Ebbinghaus. Memory: A contribution to experimental psychology. *Annals of Neurosciences*, 1885.

[Fallahzadeh and Ghasemzadeh, 2017] R. Fallahzadeh and H. Ghasemzadeh. Personalization without user interruption: Boosting activity recognition in new subjects using unlabeled data. In *2017 ACM/IEEE 8th International Conference on Cyber-Physical Systems (ICCPS)*, April 2017.

[Hoi *et al.*, 2006] Steven C. H. Hoi, Rong Jin, Jianke Zhu, and Michael R. Lyu. Batch mode active learning and its application to medical image classification. In *23rd International Conference on Machine Learning*, Pittsburgh, PA, 2006.

[Huang *et al.*, 2014] Sheng-Jun Huang, Rong Jin, and Zhi-Hua Zhou. Active learning by querying informative and representative examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10), October 2014.

[Jun and Nowak, 2016] Kwang-Sung Jun and Robert Nowak. Graph-based active learning: A new look at expected error minimization. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Washington, DC, USA, December 2016.

[Ma *et al.*, 2019] Yuchao Ma, Zhila Esna Ashari, Mahdi Pedram, Navid Amini, Daniel Tarquinio, Kouros Nouri-Mahdavi, Mohammad Pourhomayoun, Robert D Catena, and Hassan Ghasemzadeh. Cyclepro: A robust framework for domain-agnostic gait cycle detection. *IEEE Sensors Journal*, 19(10):3751–3762, 2019.

[Murukannaiah and Singh, 2015] Pradeep K. Murukannaiah and Munindar P. Singh. Platys: An active learning framework for place-aware application development and its evaluation. *ACM Transactions on Software Engineering and Methodology*, 24(3), May 2015.

[Palumboa *et al.*, 2016] Filippo Palumboa, Claudio Gallicchiob, Rita Pucciband, and Alessio Micheli. Human activity recognition using multisensor data fusion based on reservoir computing. *Journal of Ambient Intelligence and Smart Environments*, 8(2):87–107, October 2016.

[Pisinger, 2005] David Pisinger. Where are the hard knapsack problems? *Computers & Operations Research*, 32(9), 2005.

[Redmond *et al.*, 2014] SJ Redmond, NH Lovell, GZ Yang, A Horsch, P Lukowicz, L Murrugarra, and M Marschollek. What does big data mean for wearable sensor systems? *Yearbook of medical informatics*, 23(01):135–142, 2014.

[Sabato and Hess, 2016] Sivan Sabato and Tom Hess. Interactive algorithms: from pool to stream. In *Proceedings of the 29th Annual Conference on Learning Theory (COLT), JMLR Workshop and Conference Proceedings*, New-York City, USA, June 2016.

[Sener and Savarese, 2018] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *ICLR 2018*, Vancouver, Canada, May 2018.

[Settles, 2012] Burr Settles. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1), 2012.

[Shen and Zhai, 2005] Xuehua Shen and ChengXiang Zhai. Active feedback in ad hoc information retrieval. In *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brazil, August 2005.

[Stikic and Schiele, 2009] Maja Stikic and Bernt Schiele. Activity recognition from sparsely labeled data using multi-instance learning. In *Proceedings of 4th International Symposium on Location- and Context-Awareness*, Tokyo, Japan, May 2009.

[Vijayanarasimhan *et al.*, 2016] Sudheendra Vijayanarasimhan, Prateek Jain, and Kristen Grauman. Far-sighted active learning on a budget for image and video recognition. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, February 2016.

[Xu *et al.*, 2007] Zuobing Xu, Ram Akella, and Yi Zhang. Incorporating diversity and density in active learning for relevance feedback. In *Proceedings of 29th European Conference on IR Research*, Rome, Italy, April 2007.