Context-Driven Online Learning for Activity Classification in Wireless Health

Jie Xu, James Y. Xu, Linqi Song, Gregory J. Pottie and Mihaela van der Schaar
Electrical Engineering Department
University of California, Los Angeles
Los Angeles, CA 90095
Email: {jiexu, jyxu, songlinqi}@ucla.edu, {pottie, mihaela}@ee.ucla.edu

***Abstract*—Enabling accurate and low-cost classification of a range of motion activities is of significant importance for wireless health through body worn inertial sensors and smartphones, due to the need by healthcare and fitness professonals to monitor exercises for quality and compliance. This paper proposes a novel contextual multi-armed bandits approach for large-scale activity classification. The proposed method is able to address the unique challenges arising from scaling, lack of training data and adaptation by melding context augmentation and continuous online learning into traditional activity classification. We rigorously characterize the performance of the proposed learning algorithm and prove that the learning regret (i.e. reward loss) is sublinear in time, thereby ensuring fast convergence to the optimal reward as well as providing short-term performance guarantees. Our experiments show that the proposed algorithm outperforms existing algorithms in terms of both providing higher classification accuracy as well as lower energy consumption.**

I. Introduction

One of the biggest problems faced by many nations this century is the staggering cost of health care and the ever growing number of people suffering from disabilities requiring continued rehabilitation in community. In just the United States, stroke alone produces around 650,000 survivors each year, most of whom require physical rehabilitation long after discharge from a hospital [1]. Across all diseases, roughly 40% of the elderly population experience one or more forms of disabilities requiring rehabilitation and this number grows at a significant rate [2][3].

The advent of microelectronics and powerful mobile devices has created a number of new opportunities in wireless health that allow us to address this serious challenge. In particular, inexpensive and pervasive remote activity monitoring through body worn inertial sensors and smartphones is now a key strategy to provide monitoring and ensure compliance for physical exercises prescribed to patients, regardless of their ability to access health care professionals in person [4][5]. While much work done so far has focused on the accurate detection of physical activities [6]-[11], many challenges remain that prevent the technique from being widely adopted: 1) Domain experts such as clinicians come from diverse backgrounds with unique sets of activities of interest. As the number of potential motions increase, traditional classifiers suffer from degraded performance and reliability; 2) Most activity classifiers' performance directly relates to the amount of training data the intended individual is able to provide. Our in-field experience in conducting large scale international trials [11] has demonstrated that this data is very difficult, if not impossible to obtain due to lack of training from healthcare professionals and patient's inability to physically perform the activities for the required training time; 3) State of the art activity classifiers are often deployed without the means to retrain and adapt to a user's improving situation, significantly reducing their usefulness after a few months.

In this paper, we present a novel contextual multi-armed bandits (MAB) approach that enables efficient and large-scale activity classification. This method is able to address the challenges highlighted above by melding context augmentation and continuous online learning into traditional activity classification. The context information in this paper is defined as the side information of physical activities, such as the location where the activity takes place, the user profile such as his/her age, gender weight etc. First, the use of context effectively subdivides the potential activity search space at any given time, making large activity classification models manageable and provides additional distinguishing features. Second, the use of continuous online learning allows the system to start with classifiers that are trained using generic data sets (easy to obtain) and gradually adapt to individual users. In this way, the classifiers can also adapt themselves to a user's improving situation by selecting more suitable models and boot-strapping the model with newer data. Finally, the analytical nature of MAB allows us to rigorously characterize the performance bounds of the models in use. We prove the regret (i.e. reward loss due to learning) incurred by our algorithm is sublinear in time, thereby ensuring fast convergence to the optimal reward as well as providing short-term performance guarantees.

The rest of this paper is organized as follows. Section II discusses related works. Section III describes the system model and formulates the problem. Second IV proposes the online learning algorithm for activity classification and bounds its learning regret. Section V provides simulation results using real-world data. Section VI concludes the paper.

II. Related Works

The benefits of activity monitoring through sensors have been demonstrated in many existing works [7]-[11]. For instance, one system for measuring home-based physical rehabilitation has been described in [7]. Using a signature detection algorithm and accelerometer's signal vector magnitude as a feature, the system detects if a user has performed a set of rehabilitation exercises accurately, and provides appropriate feedback. In [8], a human activity classification system was developed for promoting exercises in an effort to reduce injuries. The system uses multiple on body accelerometers, and a large number of binary classifiers each trained to recognize specific activities. A series of optimizations link the individual classifiers to produce a final output. Most methods confront the challenge of classifying a specific motion among many possibilities at any observation time [11]. As the number of potential motions increase, the classifier model complexity increases and classification performance and reliability are degraded. In addition, these systems do not address the issue of rapid adaptation to the demands of large heterogeneous user communities, where different activities are of interest, requiring separate models, classification methods and features. This paper develops online algorithms for learning the best activity classifiers by exploiting the contextual information of the unclassified motions, thereby improving the classification performance. The proposed algorithm is developed under the contextual multi-armed bandits (MAB) framework. Previously, MAB methods were applied to solve problems in clinical trials [12][13], multi-user communication networks [14], web advertising [15], recommender systems [16][17] and stream mining systems [18][19]. A key advantage of MAB methods as compared to other online learning methods is that they can provide a bound on the convergence speed as well as a bound on the loss due to learning compared to an oracle solution which requires knowledge of the stochastic model of the system, which is named regret. To the authors' best knowledge, it is the first time that MAB methods are applied to solve activity classification problems in eHealth systems.

III. System Model

A. System Architecture

We consider a wireless activity monitoring system shown in Figure 1. Healthcare providers prescribe individualized exercise plans dependent on a subject's needs and monitor them for quality and compliance. At the end-user side, a set of Bluetooth sensors is needed with a smart mobile device to provide data to a backend server through wifi/cellular, where context and activity classification decisions are made. The returned results can be consumed by third party applications. In this architecture, the server components mainly include a Context Classification Module (CCM) and an Activity Classification Module (ACM). In this paper, we focus on the activity classification problem using the context information provided by the CCM.

B. Activity Classification Module

We assume that time is divided into discrete slots $t = 1, 2, \ldots$. At the beginning of each time slot t, the ACM on the server receives one activity classification request along with the context information from the wireless sensors of a user. The context information is provided by the CCM and can include information about the location in which this

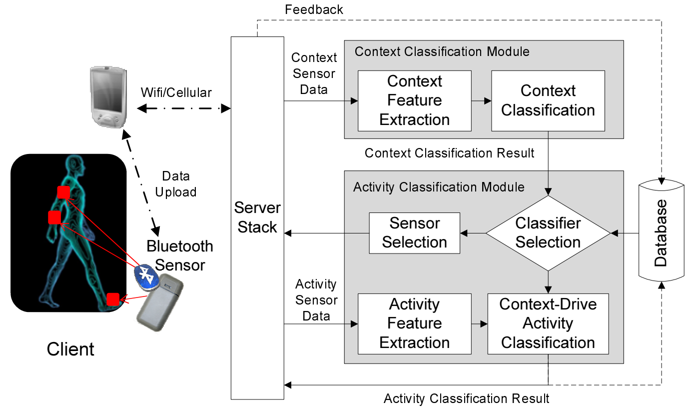

Fig. 1. System architecture.

unclassified activity takes place as well as information about the user profile such as age, gender and weight etc. We abstract the context information at time t using the notation $\theta^t \in \Theta$ with Θ being a d-dimensional metric space.

The ACM maintains a finite set of K activity classifiers $\mathcal{F} = \{f_1, ..., f_K\}$. Each classifier f_k takes input using the activity data from a specific set of sensors with respect to f_k and outputs a classification result. That is, f_k is a function $f_k : \mathcal{X}_{f_k} \to \mathcal{Y}$ where $\mathcal{X}_{f_k}$ is the sensor data value space with respect to f_k-specific sensor set and $\mathcal{Y}$ is the activity value space. Because different classifiers require the input data from different sets of sensors, invoking them incurs different costs such as sensing and wireless transmission energy consumption.

At each time slot t, given the context information θ^t, the ACM chooses one classifier f^t from the set $\mathcal{F}$ to perform the activity classification. Depending on the model selected, a specific set of sensors is invoked, incurring a cost of c^t which is reported to the ACM at the end of the time slot. Once the sensor data $x^t \in \mathcal{X}_{f^t}$ is collected and provided to the ACM, the ACM uses f^t to classify the activity. Let the classification result be $y^t \in \mathcal{Y}$. We assume that at the end of time slot t, the ground-truth label of activity $\hat{y}^t \in \mathcal{Y}$ is revealed. This label may be provided by the end-user himself/herself or by the physicians. It can be erroneous, provided occasionally and/or with delay. For the sake of analysis simplicity, we will assume that the label is provided immediately at the end of each period. Let $a^t = I(y^t = \hat{y}^t)$ be the classification accuracy where $I(\cdot)$ is the indicator function. Thus, the problem under consideration features a supervised online learning problem: after selecting a classifier f_t, a realized classification reward r^t which jointly takes into account the classification accuracy a^t and the associated classification cost c^t (e.g. sensing and wireless transmission energy consumption) is revealed to the ACM. For example, the classification reward can be a linear combination of accuracy and cost, i.e. $r^t = a^t - \gamma c^t$ where γ is a trade-off parameter.

C. Classification Reward and Learning Regret

For each context $\theta \in \Theta$, selecting a classifier f yields an (unknown) *expected* classification reward $\mu_\theta(f) = Er_\theta(f)$. Notice that for each specific activity classification request with context information θ, the realized reward $r_\theta(f)$ by selecting the classifier f is a random variable drawn from an unknown distribution with mean $\mu_\theta(f)$ which is also unknown. Let $f^*(\theta) := \arg\max_{f \in \mathcal{F}} \mu_\theta(f)$ and $\mu^*_\theta := \mu_\theta(f^*(\theta))$. We call $f^*(\theta)$ the *best oracle* classifier for context θ. Notice that the best oracle classifiers are not known by the ACM but instead need to be learned.

A learning algorithm σ selects for the context information θ^t at each time slot t a classifier $\sigma^t \in \mathcal{F}$ to use. The regret (learning loss) of a learning algorithm used by the ACM with respect to the oracle benchmark by time T is given by

$$\text{Reg}(T) = \sum_{t=1}^{T} \mu^*_{\theta^t} - E\left[\sum_{t=1}^{T} r^t_{\theta^t}(\sigma^t)\right] \tag{1}$$

Regret gives the convergence rate of the total expected classification reward of the learning algorithm to the value of the optimal solution. Any algorithm whose regret is sublinear, i.e. $\text{Reg}(T) = O(T^\gamma)$ such that $\gamma < 1$, will converge to the optimal solution in terms of the average reward. The goal of the ACM is to minimize its regret, which is equivalent to maximizing the total reward by time T.

In the next section, we will propose an efficient learning algorithm that learns the optimal classifiers with sublinear regret bounds. To enable rigorous regret analysis, we will make the following widely adopted technical assumption.

Assumption. *(Lipschitz) For each* $f \in \mathcal{F}$*, there exists* $L > 0, \alpha > 0$ *such that for all* $\theta, \theta' \in \Theta$*, we have* $|\mu_\theta(f) - \mu_{\theta'}(f)| \leq L\|\theta, \theta'\|^\alpha_\Theta$.

The above assumption states that if context information is similar, then the expected reward by selecting the same classifier is also similar.

IV. Context-Driven Activity Classification

The basic idea of our online learning algorithm is as follows. The algorithm alternates between two phases over time. In the exploration phases, different classifiers are explored to learn their expected classification reward. In the exploitation phases, the classifier with the best estimated classification reward is selected in order to maximize the classification reward. This learning problem would be simple if there was no context information. However, without using the context information the performance of the learning algorithm can be poor because the best oracle classifiers can be very different for different context information. However, since the context space Θ can be very large and even continuous, learning the best oracle classifier for each individual context $\theta \in \Theta$ is extremely difficult, if not impossible, for a finite number T of activity classification requests. To overcome this obstacle, our learning algorithm will exploit the similarity information of contexts, adaptively and dynamically partition the context space into smaller subspaces and learn the best oracle classifier within each subspace.

A. Algorithm Description

In this subsection, we describe the proposed online learning algorithm for activity classification. For analysis simplicity, we normalize the context space to be $\Theta = [0, 1]^d$. The following notions are important for the proposed algorithm.

- **Context Space Partition.** By uniformly partitioning the context space on each dimension by l, we create 2^{ld} context subspaces, each of which is a d-dimensional hypercube with side length being 2^{-l}. We call this partition a level l partition $\mathcal{P}_l$ and clearly $|\mathcal{P}_l| = 2^{ld}$. Note that $\mathcal{P}_0$ contains only a single hypercube which is the entire context space Θ. Let $\mathcal{P} := \cup_{l=0}^{\infty}\mathcal{P}_l$ denote the set of all possible such subspaces.
- **Active Context Subspace.** In each time slot, the algorithm keeps a set of mutually exclusive context subspaces that cover the entire context space. We call these subspaces the *active* subspaces, and denote the set of active subspaces at time t by $\mathcal{A}^t$. Clearly we have $\cup_{s \in \mathcal{A}^t} = \Theta, \forall t$.
- **Activation, Partitioning and Deactivation.** Once a subspace $C \in \mathcal{P}$ is activated, we maintain a counter N_C that records the number of times that context arrives to C. A level l subspace C will stay active until the first time t such that $N_C \geq A2^{pl}$ where $p > 0$ and $A > 0$ are algorithm design parameters. At this point, the level l subspace C is further partitioned into 2^d smaller $l+1$ subspaces that constitutes C. Then C becomes inactivate and these smaller subspaces become active and 2^d new counters are created.
- **Reward Estimates.** For each active context subspace C, the algorithm maintains the estimated rewards $\bar{r}_C(f)$ for all classifiers for the context arrivals to this subspace.
- **Counters.** For each active context subspace C, the algorithm maintains several counters. The first counter M_C records the number of context arrivals to C which is used for context subspace partitioning. For each subspace C, the algorithm also maintains for each classifier f a counter $N_C(f)$ that records the number of times f is selected to classify the request. Clearly, $M_C = \sum_{f \in \mathcal{F}} N_C(f)$ at any moment in time.
- **Control Function.** The algorithm uses a control function $D(t)$ which takes time as the input and outputs a real number. The control function has the form of $D(t) = t^z \ln t$.

The algorithm works as follows. When an activity classification request with context information θ^t comes at time t, the algorithm checks to which active subspace $C \in \mathcal{A}^t$ it belongs. Then it investigates counters $N_C(f)$ for all classifiers to see if there exists any classifier f such that $N_C(f) \leq D(t)$. There are two cases:

- If there exists such a classifier f, then the algorithm selects this classifier for the current request, i.e. $\sigma^t = f$. (**Exploration**)
- If there does not exist such a classifier, then the algorithm selects the classifier with the highest reward estimate $\sigma^t = \arg\max_{f \in \mathcal{F}} \bar{r}_C(f)$. (**Exploitation**)

At the end of time t, the actual classification reward is revealed to be r^t which then is used to update the reward estimate $\bar{r}_C(f)$. If the context arrival counter for this context

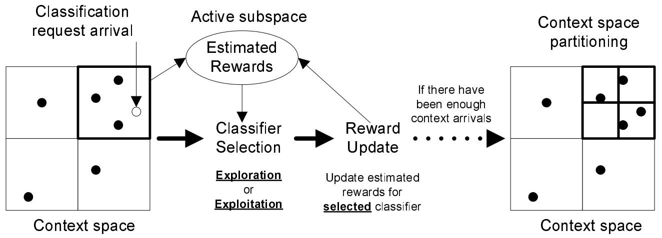

Fig. 2. Context space partitioning

subspace N_C exceeds $A2^{pl}$, then the context subspace is further partitioned. The formal description is presented in Algorithm 1. A pictorial illustration is provided in Figure 2.

Algorithm 1 Online Learning for Activity Classification

```
Initialize A_1 = P_0, M_Θ = 0, r̄_Θ(f) = 0, ∀f ∈ F.
for each activity classification request at time t do
    Determine C ∈ A^t such that θ ∈ C.
    if ∃f ∈ F such that N_C(f) < D(t) then
        Select σ^t = f. (Exploration)
    else
        Select σ^t = arg max_{f∈F} r̄_C(f). (Exploitation)
    end if
    Set N_C(σ^t) ← N_C(f^t) + 1
    (The activity classification reward r^t is revealed.)
    Update r̄_C(σ^t).
    Set M_C ← M_C + 1.
    if M_C ≥ A2^{pl} then
        Set A^{t+1} = (A^t \ C) ∪ P_{l+1}(C)
    end if
end for
```

B. Learning Regret Analysis

In this subsection, we analyze the regret of the proposed learning algorithm for activity classification.

The following notations will be useful for our analysis. Let $\Delta = \max_{\theta\in\Theta}\{\mu^*_\theta - \min_{f\in\mathcal{F}}\mu_\theta(f)\}$ be the maximized reward difference between the best oracle classifier and a non-optimal classifier. Without loss of generality, we normalize $\Delta = 1$. Let $\mathcal{E}^t_C(f)$ be the set of rewards collected by selecting classifier f by time t for active subspace C. For each subspace C let $f^*(C)$ be the classifier which is optimal for the center context of that subspace, and let $\bar{\mu}_C(f) := \sup_{\theta\in C}\mu_\theta(f)$ and $\underline{\mu}_C(f) := \inf_{\theta\in C}\mu_\theta(f)$. For a level l subspace C, we define the set of *sub-optimal* classifiers to be

$$\mathcal{L}_C(B) := \{f : \underline{\mu}_C(f^*) - \bar{\mu}_C(f) > BLd^{\alpha/2}2^{-l\alpha}\} \quad (2)$$

where B is a constant. Finally, let $\beta_a := \sum_{t=1}^{\infty} 1/t^a$.

We decompose the regret of learning into three parts

$$\text{Reg}(T) = \text{Reg}_e(T) + \text{Reg}_s(T) + \text{Reg}_n(T) \quad (3)$$

where $\text{Reg}_e(T)$ is the regret due to exploration, $\text{Reg}_s(T)$ is the regret due to *sub-optimal* classifier selections in exploitation and $\text{Reg}_n(T)$ is the regret due to *near-optimal* classifier selections in exploitation. The following series of lemmas bound each of these terms separately.

We start with a simple lemma which gives an upper bound on the highest level subspace that is active at any time t.

Lemma 1. *All active subspaces in $\mathcal{A}^t$ at time t have a level of at most $(\log_2 t)/p + 1$.*

Proof: Let $l + 1$ be the highest level. According to the partitioning process, we must have $\sum_{j=1}^{l} A2^{pj} < k$, otherwise the highest level will be less than $l + 1$. From this, for $t > A$, we have $l < \log_2(t)/p$. ■

The next three lemmas bound the regret for any level l context subspace.

Lemma 2. *If $D(t) = t^z \ln t$, then for any level l subspace, the regret due to exploration by time t is bounded above by $K(t^z \ln t + 1)$.*

Proof: Time slot t is an exploration slot if and only if there exists f such that $N^t_C(f) \le D(t)$. Therefore, up to time T, there can be at most $t^z \ln t + 1$ exploration time slots for each classifier f. Since there are a total number of K classifiers, the number of exploration slots is upper bounded by $K(t^z \ln t + 1)$. ■

Lemma 3. *Let $B = \frac{2}{Ld^{\alpha/2}2^{-\alpha}} + 2$. If $p > 0, 2\alpha p \le z < 1$, $D(t) = t^z \ln t$, then for any level l subspace C, the regret due to choosing sub-optimal classifiers in exploitation steps is bounded by $2K\beta_2$.*

Proof: Let Ω denote the space of all possible outcomes, and w be a sample path. The event that the algorithm exploits in C at time t is given by

$$\mathcal{W}^t_C := \{w : N_C(f) > D(t), \forall f; \theta^t \in C; C \in \mathcal{A}^t\} \quad (4)$$

We will bound the probability that the algorithm chooses a sub-optimal classifier in an exploitation step in subspace C, and then bound the expected number of times a sub-optimal classifier is chosen by the algorithm. Recall that the loss in every slot is at most $\Delta = 1$. Let $\mathcal{V}^t_C(f)$ be the event that a sub-optimal classifier f is chosen. Then

$$\text{Reg}_{C,s}(T) \le \sum_{t=1}^{T} \sum_{f\in\mathcal{L}_C(B)} P(\mathcal{V}^t_C(f), \mathcal{W}^t_C) \quad (5)$$

For any classifier f, we have

$$\{\mathcal{V}^t_C(f), \mathcal{W}^t_C\} \quad (6)$$
$$\subset \{\bar{r}_C(f) \ge \bar{\mu}_C(f) + H_t, \mathcal{W}^t_C\} \quad (7)$$
$$\cup \{\bar{r}_C(f^*) \le \underline{\mu}_C(f^*) - H_t, \mathcal{W}^t_C\} \quad (8)$$
$$\cup \{\bar{r}_C(f) \ge \bar{r}_C(f^*), \bar{r}_C(f) < \bar{\mu}_C(f) + H_t, \quad (9)$$
$$\bar{r}_C(f^*) > \underline{\mu}_C(f^*) - H_t, \mathcal{W}^t_C\} \quad (10)$$

for some $H_t > 0$. This implies

$$\begin{aligned}
&P(\mathcal{V}_C^t(f), \mathcal{W}_C^t) \\
\leq &P(\bar{r}_C^{best}(N_C(f)) \geq \bar{\mu}_C(f) + H_t + Ld^{\alpha/2}2^{-l\alpha}, \mathcal{W}_C^t) \\
+ &P(\bar{r}_C^{worst}(N_C(f^*)) \leq \underline{\mu}_C(f^*) - H_t - Ld^{\alpha/2}2^{-l\alpha}, \mathcal{W}_C^t) \\
+ &P(\bar{r}_C^{best}(N_C(f)) \geq \bar{r}_C^{worst}(M_C(f^*)), \\
&\quad \bar{r}_C^{best}(N_C(f)) < \bar{\mu}_C(f) + H_t, \\
&\quad \bar{r}_C^{worst}(N_C(f^*)) > \underline{\mu}_C(f^*) - H_t, \mathcal{W}_C^t)
\end{aligned}$$

Consider the last term in the above equation. In order to make it 0, we need $2H_t \leq (B-2)Ld^{\alpha/2}2^{-l\alpha}$. This holds when $2H_t \leq (B-2)Ld^{\alpha/2}2^{-\alpha}t^{-\alpha/p}$. Therefore, for $H_t = t^{-z/2}, z \geq 2\alpha/p$ and $B = \frac{2}{Ld^{\alpha/2}2^{-\alpha}} + 2$, the last term is 0. Then, by using a Chernoff- bound, for any $f \in \mathcal{L}_C(B)$, since on the event $\mathcal{W}_C^t$, $N_C(f) \geq t^z \ln t$, the first two terms in the last line in the above equation are both bounded by

$$e^{-2(H_t)^2 t^z \ln t} \leq t^{-2} \tag{11}$$

Therefore $\text{Reg}_{C,s}(T) \leq \sum_{t=1}^{T} \sum_{f \in \in \mathcal{L}_C(B)} t^{-2} \leq K\beta_2$. ■

Lemma 4. *Let $B = \frac{2}{Ld^{\alpha/2}2^{-\alpha}} + 2$. If $p > 0, 2\alpha p \leq z < 1$, $D(t) = t^z \ln t$, then for any level l subspace C, the regret due to choosing near-optimal classifiers in exploitation slots is bounded above by $2ABLd^{\alpha/2}2^{(p-\alpha)l}$.*

Proof: The one-slot regret of any near-optimal classifier f is bounded by $2BLd^{\alpha/2}2^{-l\alpha}$. Since C remains active for at most $A2^{pl}$ context arrivals, we obtain the desired bound. ■

In order to obtain the regret bound of the proposed online activity classification algorithm, we need to consider how many subspaces of each level is formed up to time T. The number of such subspaces explicitly depends on the context information arrival process. Therefore, we investigate the regret for different context arrival scenarios.

Definition. *We call the context arrival process* ***the worst case arrival process*** *if it is uniformly distributed inside the context space, with minimum distance between any two context samples being $T^{-1/d}$, and* ***the best case arrival process*** *if $x^t \in C, \forall t$ for some level $\lceil \log_2(T)/p \rceil + 1$ subspace C.*

The following theorems determine the finite time, uniform regret bounds for the online activity classification algorithm.

Theorem 1. *For the worst case arrival process, Reg(T) = $O(T^{\frac{d+\alpha/2+\sqrt{9\alpha^2+8\alpha d}/2}{d+3\alpha/2+\sqrt{9\alpha^2+8\alpha d}/2}})$ by choosing $p = \frac{3\alpha+\sqrt{9\alpha^2+8\alpha d}}{2}$ and $z = 2\alpha/p$.*

Proof: Let $B = \frac{2}{Ld^{\alpha/2}2^{-\alpha}} + 2$. We first consider the worst case. It can be shown that in the worst case the highest level subspace has level at most $1 + \log_{2p+d} T$. The total number of subspaces is bounded by

$$\sum_{l=0}^{1+\log_{2p+d} T} 2^{dl} \leq 2^{2d} T^{\frac{d}{d+p}} \tag{12}$$

According to Lemma 4, the regret from choosing a near optimal classifier is

$$\text{Reg}_n(T) \leq 2ABLd^{\alpha/2} \sum_{l=0}^{1+\log_{2p+d} T} 2^{(p-\alpha)l} \tag{13}$$

$$\leq 2ABLd^{\alpha/2}2^{2(d+p-\alpha)}T^{\frac{d+p-\alpha}{d+p}} \tag{14}$$

Hence, $\text{Reg}_n(T)$ is on the order of $T^{\frac{d+p-\alpha}{d+p}}$. Moreover, since the number of activated subspaces is on the order of $O(T^{\frac{d}{d+p}})$, according to Lemma 2, $\text{Reg}_e(T)$ is on the order of $O(T^{\frac{d}{d+p}+z} \ln T)$ and according to Lemma 3, $\text{Reg}_s(T)$ is on the order of $O(T^{\frac{d}{d+p}+z})$, for $z \geq \frac{2\alpha}{p}$. These three parts of the regret are balanced when $z = 2\alpha/p$ and $\frac{d+p-\alpha}{d+p} = \frac{d}{d+p} + z$. Solving for p we get

$$p = \frac{3\alpha + \sqrt{9\alpha^2 + 8\alpha d}}{2} \tag{15}$$

Therefore, the regret is $\text{Reg}(T) = O(T^{\frac{d+\alpha/2+\sqrt{9\alpha^2+8\alpha d}/2}{d+3\alpha/2+\sqrt{9\alpha^2+8\alpha d}/2}})$. ■

Theorem 2. *For the best case arrival process, Reg(T) = $O(T^{2/3})$ by choosing $p = 3\alpha$ and $z = 2\alpha/p$.*

Proof: Now we consider the best case, the number of activated subspaces is upper bounded by $\log_2 T/p + 1$, and by the property of context arrivals all activated subspaces have different levels. We calculate the regret from choosing near optimal classifiers as

$$\text{Reg}_n(T) \leq 2ABLd^{\alpha/2} \sum_{l=0}^{1+\log_2 T/p} 2^{p-\alpha}l \tag{16}$$

$$\leq 2ABLd^{\alpha/2} \frac{2^{2(p-\alpha)}}{2^{p-\alpha}} T^{\frac{p-\alpha}{p}} \tag{17}$$

The other regret parts are the same as the worst case. These three parts are balanced by setting $z = 2\alpha/p$, $p = 3\alpha$. Substituting these parameters we obtain $\text{Reg}(T) = O(T^{2/3})$. ■

The regret bounds proved in Theorem 1 and 2 are sublinear in time T which guarantee convergence in terms of the average classification rewards, i.e. $\lim_{T \in \infty} \text{Reg}(T)/T = 0$. Thus our online learning algorithm makes the optimal classification as sufficient classification requests have arrived. More importantly, the regret bound tells how much reward would have be lost by running our learning algorithm for any finite time T. Hence, it also provides a rigorous characterization on the short-term learning performance.

C. Extensions

In the above analysis, we assumed that the true labels of the user activity are revealed immediately at the end of every slot. In practice, labels can be missing or received with a delay. In the case of delayed labels, the proposed learning algorithm can be easily modified to update the expected reward of classifiers as soon as the label is received. In this way, if the delay is finite, the regret orders given by Theorem 1 and 2 are not affected. Only a constant term, which is a function of the maximum delay is added to the regret. In the case of missing

	f_1	f_2	f_3	f_4
Upper right arm sensor	X	X		
Lower right arm sensor	X			X
Lower right leg sensor	X	X	X	X

TABLE I
SENSORS REQUIRED FOR DIFFERENT CLASSIFIERS

labels, it can be shown analytically that, if the probability of not receiving the label at any time t is q, the regret orders will be scale by $1/(1-q)$ but the sublinearity with respect to time is not affected either.

Next, we assess the computation and memory requirements of the proposed algorithm. For each active subspace $C \in \mathcal{A}^t$, the algorithm needs to keep the sample mean reward estimates of K classifiers. A level l active subspace becomes inactive if the context arrivals to that subspace exceeds $A2^{pl}$. For the worst case arrival, the number of active subspace at time T is upper bounded by $O(\frac{d}{(p+d)A}T)$. Thus, the memory requirement is $O(K\frac{d}{(p+d)A}T)$. For the best case arrival, the number of active subspace at time T is upper bounded by $O(\frac{2^d}{p}\log_2 T)$. Thus, the memory requirement is $O(K\frac{2^d}{p}\log_2 T)$. However, the algorithm can be modified so that the available memory provides an an upper bound the deepest level of the context subspace.

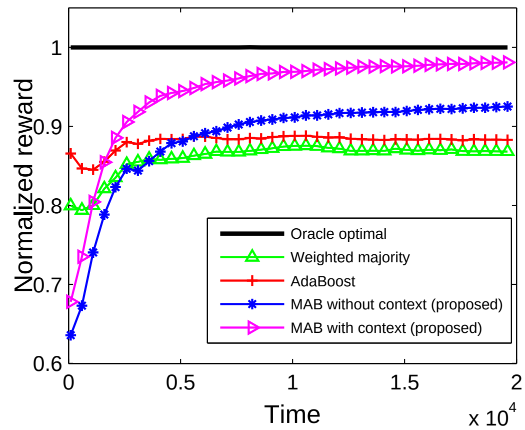

Fig. 3. Normalized classification reward comparison for $\gamma = 0.1$.

	Running	Walking Around	Walking Normal
MAB w. context (proposed)	0.98	0.98	0.97
Weighted Majority	0.84	0.77	0.69
AdaBoost	0.85	0.80	0.70

TABLE II
NORMALIZED CLASSIFICATION REWARD FOR INDIVIDUAL ACTIVITIES ($\gamma = 0.1$).

V. SIMULATIONS

Our experiments are performed using real-world sensor data collected from end-users. The end-user component is a physical package containing four IMUs with Velcro attachments, a Nexus 7 tablet and associated applications. A more detailed description of our system deployment can be found in [6].

The server continuously receives classification requests from end-users through wifi/cellular. For each request, the CCM first detects the context information associated with the request. Then the ACM selects a classifier to perform the classification using this context information. Since all experiments are using data from the same location, the context information which we use are the gender and age of the end-user. We consider three possible activities: "Running", "Walking Normal" and "Walking Around". Four pre-trained classifiers are implemented, each requiring different sensors to be activated and using different features as the input (see Table I). The proposed learning algorithms (with and without using the context information) are compared against two benchmark solutions:

- **Weighted majority (WM)**. The second benchmark is the widely adopted weighted majority algorithm [20].
- **AdaBoost**. The third benchmark implements an online version of the famous AdaBoost algorithm [21].

We note that for both benchmarks, since all classifiers are used, all sensors need to be activated for all requests.

In all experiments, the reward function has the form of $r^t = a^t - \gamma c^t$ where a^t is the prediction accuracy and c^t depends on the number of sensors activated for classification. Figure 3 shows the obtained average per period reward (normalized to the optimal reward obtained by the oracle solution) over time for $\gamma = 0.1$. The proposed learning algorithms significantly outperform both benchmark solutions. Compared with WM and Adaboost, the proposed learning algorithms gain 15% more rewards. Since the context information used in our experiments is limited (i.e. only age and gender), the performance improvement by using the context information is moderate. However, much more performance improvement can be expected when more context information (e.g. location, user weight) is available. In Table II, we take a closer look at the classification results for each individual activities ($\gamma = 0.1$). WM and AdaBoost perform moderately well for the "Running" activity but the rewards for the two similar activities, "Walking Around" and "Walking Normal", are very low. Instead, the proposed algorithm is able to achieve high classification rewards for all three activities.

Finally, we investigate the trade-off between energy consumption (i.e. classification cost) and classification accuracy. Figure 4 illustrates the accuracy and energy consumption trade-off curve of the proposed algorithm. The energy consumption is normalized to the maximum power consumption when all sensors are activated. Note that WM and AdaBoost use all sensors for all requests and hence, they are not able to make trade-off between energy consumption and accuracy. As can be seen from the figure, higher accuracy can be obtained at a cost of higher energy consumption for all three activities as well the the overall performance.

VI. CONCLUSIONS

In this paper, we proposed a novel online learning method for activity classification in wireless health systems using wearable inertial sensors. We have shown by incorporating the

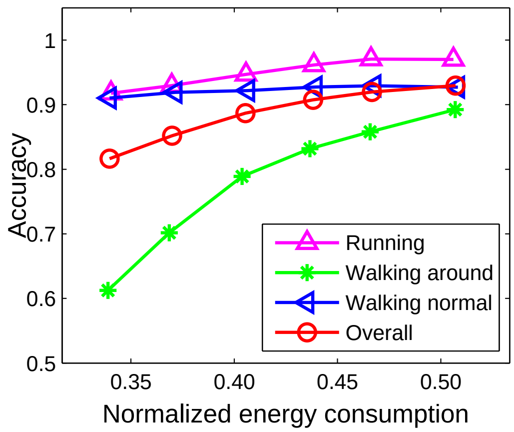

Fig. 4. Trade-off between accuracy and energy consumption (normalized to the maximum energy consumption).

contextual information, significantly better activity classification performance can be achieved than existing approaches. The proposed method does not require a priori training data, self-adapts to a user's improving situation and scales with a large number of users. We also have systematically proved sublinear regret bounds on the performance loss incurred by our algorithm due to online learning, providing both long-term and short-term performance guarantee.

References

[1] American Heart Association, "Heart disease and stroke statistic - 2010 update," Circulartion, pp. 121, 2010.

[2] A. P. Glascock, D. M. Kutzik, "Behavioral telemedicine: A new appraoch to the continuous nonintrusive monitoring of activities of daily living," Telemedicine Journal, vol. 6, no. 1, pp. 33-44, 2000.

[3] U. Varshney, "Pervasive healthcare and wireless health monitoring," Mobile Networks and Applications, vol. 12, no. 2-3, pp. 113-127, 2007.

[4] B. H. Dobkin and A. Dorsch, "The promise of mHealth: Daily activity monitoring and outcome assessment by wearable sensors," Neurorehabilitation and Neural Repair, vol. 25, pp. 788-798, 2011.

[5] M. F. Gordon, "Physical activity and exercise recommendations for stroke survivors," Stroke, pp. 1230-1240, 2004.

[6] J. Y. Xu, H.-I Chang, C. Chien, W. J. Kaiser, G. J. Pottie, "Context-driven, prescription based personal activity classification: methodology, architecture and end-to-end implementation," IEEE J. Biomedical and Health Informatics, Sept. 2013.

[7] C. Chiang, Y. Lee, C. Hsieh, C. Chan and S. J. Hsu, "Quantification of home rehabilitation excercise for the elder's physical fitness monitoring," 5th Int. Conf. on Bioinformatics and Biomedical Engineering, 2011.

[8] O. Banos, M. Damas, H. Pomares and I. Rojas, "Recognition of human physical activity based on a novel hierarchical weighted classification scheme," In Prof. of Int. Joint Conf. on Neural Networks, 2011.

[9] B. H. Dobkin, X. Xu, M. Batalin, S. Thomas, and W. J. Kaiser, "Accelerometry algorithms for activity recognition," Stroke, 2011.

[10] B. Logan and J. Healey, "Sensors to detect the activities of daily living," Engineering in Medicine and Biology, pp. 5362-5365, 2006.

[11] X. Xu, M. Batalin, B. Dobkin, W. Kaiser, "Robust hierachical system for classifcation of complex human mobility characteristics in the presence of neurological disorders," Int. Conf. on Body Sensor Networks, 2011.

[12] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," Advances in Applied Mathematics, vol. 6, pp. 4-22, 1985.

[13] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," Machine Learning, vol. 47, pp. 235-256, 2002.

[14] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," IEEE Trans. Signal Process., vol. 58, no. 11, pp. 5667-5681, 2010.

[15] A. Silvkins, "Contextual bandits with similarity information," in 24th Annual Conference on Learning Theory (COLT), 2011.

[16] L. Li, W. Chu, J. Langford and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in Proc. of the 19th international conference on world wide web, 2010.

[17] C. Tekin, S. Zhang, and M. van der Schaar, "Distributed online learning in social recommender systems," to appear in IEEE J. Sel. Topics on Signal Process, 2013.

[18] C. Tekin and M. van der Schaar, "Distributed online Big Data classification using context information," Allerton 2013.

[19] J. Xu, C. Tekin and M. van der Schaar, "Learning optimal classifier chains for real-time big data mining," Allerton 2013.

[20] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," Inf. Comput., vol. 108, no. 2, pp. 212-261, Feb. 1994.

[21] W. Fan, S. J. Stolfo, and J. Zhang, "The application of AdaBoost for distributed, scalable and on-line learning," in Proc. ACM SIGKDD, 2003.