Energy-Efficient Signal Processing in Wearable Embedded Systems: An Optimal Feature Selection Approach

Hassan Ghasemzadeh, Navid Amini, and Majid Sarrafzadeh
University of California, Los Angeles, California, United States
{hassan,amini,majid}@cs.ucla.edu

ABSTRACT

Many wearable embedded systems benefit from classification algorithms where statistical features extracted from physiological signals are mapped onto different user's states such as health status of a patient or type of activity performed by a subject. Conventionally selected features lead to rapid battery depletion in these battery-operated systems, mainly due to the absence of computing complexity criterion while selecting prominent features. In this paper, we introduce the notion of power-aware feature selection, which minimizes energy consumption of the signal processing for classification applications. Our approach takes into consideration the energy cost of individual features that are calculated in real-time. The problem is formulated using integer programming and a greedy approximation is presented to select the features in a power-efficient manner. Experimental results on thirty channels of activity data demonstrate that our approach can significantly reduce energy consumption of the computing module resulting in more than 30% energy savings while achieving 96.7% classification accuracy.

Categories and Subject Descriptors

C.3 [**Computer Systems Organization**]: Special Purpose and Application-Based Systems—*Real-time and embedded systems*; J.3 [**Computer Applications**]: Life and Medical Science—*Health*; H.1.2 [**Information Systems**]: Models and Principles—*User/Machine Systems Human information processing; Human factors.*

General Terms

Design, Algorithms, Experimentation.

Keywords

Wearable Monitoring, Embedded Signal Processing, Activity Recognition, Feature Selection, Power Optimization.

ISLPED'12, July 30–August 1, 2012, Redondo Beach, CA, USA.

1. INTRODUCTION

Recent technology advances have led to the development of different sensing, computing and communication artifacts that are becoming essential part of our daily lives. These ubiquitous platforms have proved to be effective in a number of domains ranging from medical and well-being to military and smart vehicles. A special class of these platforms is wearable monitoring where computational elements are tightly coupled with the human body. As these embedded systems continue to become more pervasive, design and development of low-power architectures that enable their sustainable realization becomes much more crucial. Low power design is even more challenging in wearable systems that are battery-operated and are known as enabling technologies for many applications such as remote patient monitoring and personalized healthcare, gaming and sports, maintenance, production and process support [1, 2].

An important aspect of low power design is to optimize power consumption of the computing modules. Wearable monitoring systems often employ embedded signal processing and machine learning blocks that use sensor data (e.g. acceleration of body joints) to extract relevant information (e.g. human movements) about their subjects.

We take special interest in classification applications, where physiological signals from human body are used to classify different states of a subject. Examples of such applications include human action recognition and fall detection using accelerometer and gyroscope sensors, and arrhythmia detection from ECG signals. In the classification process, a set of representative features, such as 'signal amplitude' and 'root mean square' power, are typically extracted from the measured signals prior to executing the classification algorithm. Feature extraction is often time consuming and can deplete the battery if an exhaustive feature set is considered. In this paper, we propose the notion of power-aware feature selection which introduces a novel approach for optimizing the power efficiency of feature extraction mechanisms. Our approach combines three criteria, namely, feature relevance, feature redundancy, and computing complexity and builds a minimum cost feature set without sacrificing the classification performance of the system.

Unfortunately, none of the feature selection techniques studied in the past takes into consideration computing complexity of the selected features, an important measure in designing wearable monitoring systems. Similarly, none of the existing power-aware schemes in embedded system design has dealt with feature selection algorithms and how the energy constraints prevent some features from being accounted for classification. Our work embodies the innovation of the notion of power-aware feature selection; we model the problem of energy optimal feature set and prove that it constitutes a computational problem that is NP-hard and finally we provide an approximation to find the appropriate minimum cost feature set. Real human motion data sets are used in order to verify the efficacy of the proposed approach.

Our contributions can be summarized as follows: 1) we introduce the notion of power-aware feature selection by adding a new design dimension, computing complexity, to the feature selection problem; 2) we propose a graph model that embodies information regarding classical feature selection, relevance and redundancy; 3) we present an Integer Linear Programming (ILP) formulation of our optimization problem and provide a greedy approximation to solve it; 4) we use real data collected from a wearable sensor network designed for movement monitoring to evaluate the effectiveness of our power optimization approach.

2. RELATED WORK

Our study spans two broad research topics, feature selection and power-aware design, that are disjointedly explored in machine learning and embedded system design fields, respectively.

In general, feature selection [3, 4] aims to find an optimal set of features from an exhaustively extracted set of features. The optimality of the solution, however, is defined by two criteria, relevance and redundancy [5]. While relevance criterion focuses on eliminating features that are irrelevant to the classification task, redundancy criterion uses inter-feature correlation measures to eliminate features with high correlation.

The tight energy constraints of battery-operated wearable monitoring systems call for the development of energy-aware signal processing methods to preserve energy [6]. Depending on the application and signals involved, it may be more useful and energy-efficient to trade off radio communications energy for processing energy, or vice versa. Dynamic sensor selection [7] which trades off classification accuracy for battery lifetime is another power saving approach. The method proposed in [8] minimizes the number of nodes necessary to obtain a given classification ratio for activity recognition.

3. PRELIMINARIES

Several basic concepts are reviewed in this section with an emphasis on the architecture of wearable monitoring platforms, their signal processing, and information extraction from physiological signals.

3.1 Wearable Monitoring Systems

A wearable monitoring system, also called body sensor network, is composed of several body-worn sensor nodes and a gateway. Each sensor node is attached to the body to sample and process physiological signals, and transmit partial results to the gateway. A node usually has several sensors for capturing different user's states (e.g. body acceleration), an embedded processor to perform limited signal processing and information extraction, and a radio for data transmissions. The gateway is a more powerful unit such as a cell phone or a PDA that performs data fusion and makes conclusions about current state of the user (e.g. 'walking', 'running', and 'sitting'). The results are further transmitted, through the Internet, to a back-end server for storage, further processing, and clinical decision support.

Figure 1: Signal processing for classification applications

The focus of this study is primarily on power optimization of the wearable sensor nodes, where stringent constrained sensor units are used to process physiological signals in real-time. Other elements (e.g. gateway and back-end server) are usually powerful in terms of computing power.

Each sensor node processes sensor readings through a chain of embedded signal processing modules, each of which is intended to extract partial information from the signal and reduces the amount of sampled data. Figure 1 illustrates a typical signal processing flow for applications targeting classifying physiological signals into user's states. In physical movement monitoring applications, readings from motion sensors such as accelerometers, magnetometers, and gyroscopes undergo signal processing to classify human actions such as 'walking', 'sit to stand', and 'jumping'. The signals that are sampled by each sensor node are first passed through a filter to reduce high frequency noise. Segmentation is intended to identify 'start' and 'end' points of the actions being classified. Feature extraction module is responsible for calculating statistical and morphological characteristics of the signal segment. Finally, a classification algorithm is utilized to determine the current state of the user (e.g. type of actions being performed by the user). Signal processing and pattern recognition require a learning phase during which the system is trained based on a training data set. During this phase, parameters of the system used in different signal processing modules are adjusted. Such parameters define the training model and will be used throughout the operation of the system. Feature selection (as specified by dashed lines in Figure 1) is only part of the learning process. The selected features, however, determine complexity of the feature extraction block during execution of the system.

3.2 Feature Relevance and Redundancy

Feature selection is a learning process aimed to find significant features for the classification application. Initially, an exhaustive set of features is extracted due to lack of domain knowledge. Many features might be redundant due to their strong correlation with other features. Yet, non-redundant features can be irrelevant to a specific classification application. Feature selection is a mature field of study in machine learning with primary selection criteria being redundancy and relevance [5].

To better present the concept of relevance and redundancy, we give a simple example adapted from [5]. Assume that the feature set $\{f_1, f_2, \dots, f_5\}$ is given, where a classification decision is made based on a binary function g as $C = g(f_1,f_2)$. Furthermore, let $f_2=\bar{f}_3$ and $f_4=\bar{f}_5$. Clearly, f_1 is a required feature as it is necessary for classification and is not redundant with respect to other features. Also, features f_4 and f_5 are redundant because they are not used to make a classification decision. We note, however, that f_2 and f_3 can be interchanged as they are equally informative. Thus, an optimal feature set is either $\{f_1,f_2\}$ or $\{f_1,f_3\}$.

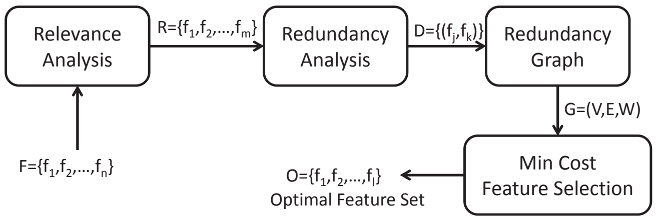

Figure 2: Proposed approach for optimal power-aware feature selection

4. PROPOSED APPROACH

Figure 2 illustrates our approach for power-aware feature selection. Given an initial feature set, F, irrelevant features are first eliminated from subsequent processing. A redundancy analysis is then performed to find features that are strongly correlated and can be substituted once power efficiency of the processing is taken into account. To this end, a graph model, called redundancy graph, is constructed based on the correlation information obtained during the redundancy analysis. Finally, the graph model is used to solve an optimization problem, called *Minimum Cost Feature Selection* (MCFS), meant for finding the optimal feature set.

4.1 Graph Model

Our relevance and redundancy analyses, which provide inputs to construct the graph model and formulate the problem, are based on the concept of symmetric uncertainty:

Definition 1 (Symmetric Uncertainty). *The symmetric uncertainty between two discrete random variables X and Y is given by:*

$$U(X,Y) = \frac{2I(X,Y)}{H(X)+H(Y)}, \tag{1}$$

where $H(X)$ and $H(Y)$ represent the entropy of random variables X and Y, respectively, and $I(X,Y)$ denotes the information gain between the two variables. $I(X,Y)$ is further defined by:

$$I(X,Y) = H(X) - H(X|Y) \tag{2}$$

The symmetric uncertainty is actually the normalized information gain and is always between 0 and 1, where U=1 means that knowing the value of either variable can completely predict the other variable, and U=0 indicates that the two variables are completely independent.

We note that the symmetric uncertainty is a measure of correlation between two random variables. Major advantage of this measure against other measures, such as correlation coefficient, is that the symmetric uncertainty can capture non-linear correlation between variables and therefore, is a safe measure for our feature analysis study.

Definition 2 (Irrelevant Feature). *Given an exhaustive set of n features F=$\{f_1, f_2, \ldots, f_n\}$ and a set of human actions $A = \{a_1, a_2, \ldots, a_h\}$ to be classified, a feature f_i is irrelevant to the classification task if*

$$min_j\,(U(f_i, a_j)) < \lambda_R, \tag{3}$$

where λ_R (relevance threshold) is a design parameter.

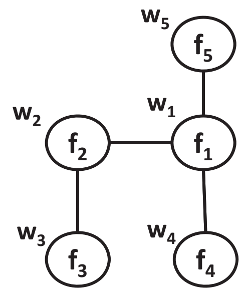

Figure 3: A motivational example for power-aware feature selection

Relevance analysis will eliminate features that are irrelevant to the action recognition. The remaining m features (m<n) are subject to redundancy analysis whose main goal is to find strongly correlated features.

Definition 3 (Strongly Correlated Features). *Two features f_i and f_k are considered to be strongly correlated if*

$$U(f_i, f_k) > \lambda_D, \tag{4}$$

where λ_D (redundancy threshold) is a design parameter.

The output of redundancy analysis is a set of feature pairs in the form of (f_i,f_k), which are strongly correlated and either of them can be eliminated according to the correlation analysis. However, these features are further examined for computing complexity using the graph model presented in the following section.

Definition 4 (Redundancy Graph). *Given m relevant features introduced by the relevance analysis and a set of feature pairs $\{f_j, f_k\}$ generated according to the redundancy analysis, an undirected graph G=(V,E,W) is called redundancy graph, where V is a set of m vertices, $V = \{u_1, u_2, \ldots, u_m\}$ associated with the m relevant features, E=$\{e_1, e_2, \ldots, e_r\}$ is the set of r feature pairs that are strongly correlated, and W=$\{w_1, w_2, \ldots, w_m\}$ is the set of weights, assigned to the vertices, denoting the computing cost associated with each feature.*

4.2 Feature Selection

We now present a simple example to motivate our idea of finding the optimal feature set using MCFS. Assume that 10 features construct our exhaustive set of features, represented by $F = \{f_1, f_2, \ldots, f_{10}\}$. Further, assume that the relevance analysis decides to eliminate five features and hence, the redundancy graph will contain five features, i.e. $R = \{f_1, f_2, f_3, f_4, f_5\}$. The redundancy graph with each vertex representing one of the five features is shown in Figure 3. Note that the processing cost attributed to each feature is represented by the weight of each vertex, e.g., w_1 is the processing cost of f_1.

Let all the weights be equal to one unit, that is $W = \{w_1, w_2, w_3, w_4, w_5\} = \{1, 1, 1, 1, 1\}$. In this case, MCFS treats all features equally and thus, the optimal feature set consists of two vertices, specifically f_1 and f_3. However, if we modify the weight set to $W = \{10, 1, 1, 1, 1\}$, MCFS gives more consideration to vertices with lower weights and accordingly, features f_4 and f_5 will be favored over f_1. In the recent scenario the optimal feature set will contain three vertices, i.e. f_4, f_5, and f_3. As such, the computation energy cost will be decreased from 11 units to 3 units.

5. MINIMUM COST FEATURE SELECTION

In this section, we present an optimization problem that finds optimal feature set taking into account relevance, redundancy and complexity criteria discussed previously. The problem takes a redundancy graph as input and outputs a subset of relevant features that are optimal in terms of computing complexity and do not exhibit any redundancy.

Problem 1. *Given a redundancy graph G=(V,E,W), the minimum cost feature selection problem is to find a subset of vertices that are not dominated by any other vertex in the graph and their total cost is minimized.*

5.1 Problem Formulation and Complexity

Assume that a_{ij} is a given binary that encodes existence of edges in the redundancy graph:

$$a_{ij} = \begin{cases} 1, & \text{if } (u_i,u_j) \in \text{V} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

and x_i is a binary variable which determines whether or not a vertex u_i is chosen as a member of the final vertex set:

$$x_i = \begin{cases} 1, & \text{if vertex } u_i \text{ is chosen} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The corresponding ILP formulation for the MCFS problem is as follows:

$$Minimize \quad \sum_{i=1}^{m} w_i x_i, \quad (7)$$

subject to:

$$\sum_{j=1}^{m} x_i a_{ij} \geq 1 \qquad \forall i \in \{1,2,\ldots,m\} \quad (8)$$

$$x_i \in \{0,1\} \quad (9)$$

The objective function in (7) is to minimize the total cost of the selected vertices (i.e. those with x_i=1). The constraint (8) guarantees that each selected vertex is adjacent to at least one more vertex and the constraint in (9) ensures that the variable x_i takes only binary values.

The MCFS problem is equivalent to the Minimum Cost Dominating Set (MCDS) problem. The MCDS problem is proved to be NP-hard by reduction from the Weighted Set Cover (WSC) problem.

Theorem 1. *The MCFS problem is NP-hard.*

Proof. Proof is eliminated for brevity. □

5.2 Greedy Approach

Our greedy algorithm for solving MCFS problem is presented in Algorithm 1. For each vertex in the redundancy graph, the algorithm first finds all adjacent vertices (V_i). It then finds the best candidate vertex to include that vertex in the final vertex set (O). The best candidate is the one with maximum value of 'cardinality of V_i divided by vertex cost w_i'. The intuition behind selecting such vertex is that it has a large number of adjacent vertices and a small cost. Finally, the algorithm adds the candidate vertex (u_i) to O and eliminates u_i and all its neighbors from V_i as well as V. The algorithm iterates until there is no more vertex in V indicating that each vertex is either chosen as a final vertex or is dominated by a final vertex.

Algorithm 1 Greedy solution for MCFS problem

Require: Redundancy graph G=(V,E,W)
Ensure: Final vertex set O
  $O = \phi$
  for all $u_i \in V$ **do**
    $V_i = \{\text{all vertices } u_j \text{ adjacent to } u_i\}$
  end for
  while $V \neq \phi$ **do**
    $V_i \leftarrow argmax_{V_i} \frac{|V_i|}{w_i}$ and $O \leftarrow O \cup u_i$
    Eliminate u_j from all V_i sets and V
  end while

Theorem 2. *Algorithm 1 achieves an* $\ln n$ *approximation to the MCFS problem.*

Proof. Proof is eliminated for brevity. □

6. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of the proposed feature analysis techniques utilizing real data collected from three human subjects using wearable motion sensors. Motion sensors were used to measure acceleration and angular velocity of six different body segments (including upper and lower body limbs) while the subjects were instructed to perform 14 transitional movements listed in tblrefmvts. The collected data were partitioned into two disjoint data sets, one for solving the proposed optimization problem and the other for measuring classification accuracy of the system on the selected features. The energy consumption of the feature extraction module was estimated on TI MSP430 microcontroller.

Table 1: Experimental movements

No.	Movement
1	Stand to Sit
2	Sit to Stand
3	Sit to Lie
4	Lie to Sit
5	Bend to Grasp
6	Rising from Bending
7	Kneeling Right
8	Rising from Kneeling
9	Look Back
10	Return from look back
11	Turn Clockwise
12	Step Forward
13	Step Backward
14	Jumping

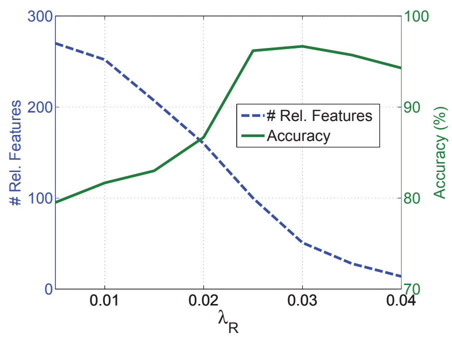

Figure 4: Number of relevant features and classification accuracy as a function of λ_R

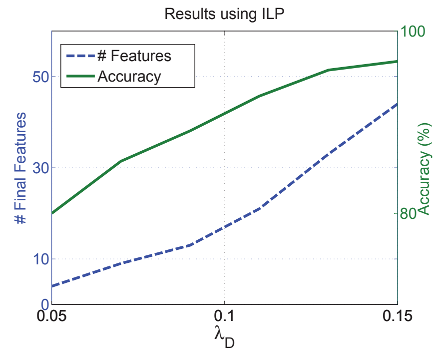

Figure 5: Optimal number of features and accuracy (using ILP solution)

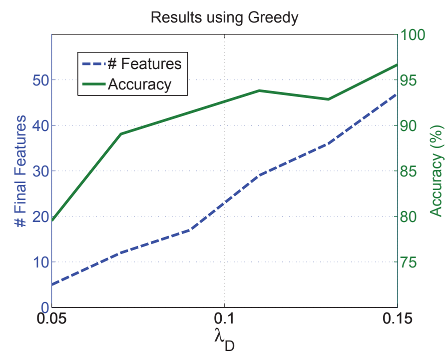

Figure 6: Number of features and accuracy for greedy algorithm

6.1 Parameter Setting

One of the parameters used to solve our optimization problem is the energy consumed for processing each feature (w_i). Energy consumption variation for different features stems from different instruction types, circuit states, and memory address modes, as well as the overall complexity of each feature. We consider TI MSP430 processor in order to find each feature's processing energy. The MSP430 is widely used in wearable monitoring applications. These systems are obviously in need of low power consumption and MSP430 nicely meets such need (594 μW power consumption on average, which yields a performance of 37 μW/MIPS).

MSP430 is a 16-bit RISC CPU that uses the Von Neumann architecture. It has 48KB of Flash memory, 10KB of RAM, and uses an 8 MHz clock. The processor benefits from a 3-stage pipeline with 16 general purpose registers. Twenty-seven instructions comprise the instruction set with 7 memory addressing modes available.

As a multiplier is a peripheral and is not implemented in every member of the MSP430 family, we utilized a method based on the Horner's approach to implement multiplication only by means of shift and add instructions. It is worthy of attention that MSP430 can perform a register shift or add in one instruction cycle.

We calculated energy consumption of feature extraction block in Figure 1 for calculating each of the 9 features listed in Table 2 using the MSP430 microcontroller, which is available on the TelosB motes used in our experiments.

6.2 Relevant Features

Figure 4 shows the number of relevant feature and classification accuracy as a function of the relevance threshold (λ_R). Clearly, the number of relevant features decreases as the threshold in (3) increases. A very small value of λ_R results in all the original 270 features being evaluated as relevant. With the complete feature set, the classification accuracy is 79.5% at the beginning. As the threshold increases, the accuracy improves as a result of irrelevant features being eliminated from the classification process. The accuracy, however, starts decreasing at $\lambda_R = 0.03$ indicating that the threshold is exceeding its optimal value and some relevant features start being eliminated from the list. Thus, we consider $\lambda_R = 0.03$ as the optimal threshold value for our relevance analysis and perform the rest of our feature analysis with this value. Note that this threshold results in 51 relevant features.

An interesting observation is that irrelevant features can significantly reduce the performance of the action recognition system. We observe that by eliminating irrelevant features, the accuracy of the classification exhibits more than 21% improvement (i.e. from 79.5% with all 270 features to 96.7% with 51 features as obtained by $\lambda_R = 0.03$ in Figure 4).

Table 2: Energy consumption of feature extraction for individual features

Feature	Energy (nJ)	Description
Amp	16386	Signal amplitude (Max - Mean)
Med	405159	Median of signal segment
Mean	8126	Mean value of signal segment
Max	8103	Maximum amplitude of signal segment
P2P	16291	Peak to peak amplitude
Var	38846	Variance of signal segment
Std	40431	Standard deviation
RMS	29705	Root mean square power
S2E	83	Start to end value

6.3 Optimal Feature Set

Figure 5 shows the number of optimal features and classification accuracy versus the redundancy threshold (λ_D) using the ILP approach. It should be noted that according to the redundancy criterion in (4), a larger value of λ_D results in less features being considered as strongly correlated and therefore, the redundancy graph will have smaller number

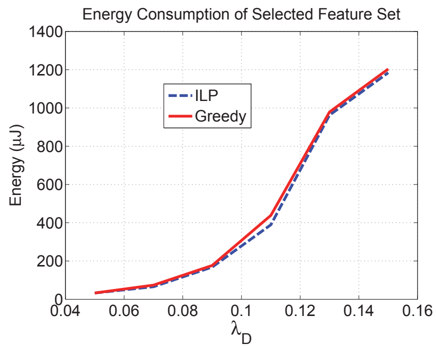

Figure 7: Energy consumption reported by optimal solution (ILP) as well as greedy approximation

of edges. With a small number of edges, more vertices need to be considered to cover all the vertices. This can result in a larger optimal set as Figure 5 shows. As the threshold increases from 0.05 to 0.15, the number of selected features grows from 4 to 44 with an average of 20.7 features. The classification accuracy ranges from 80% for λ=0.05 to 96.7% for λ=0.15. The classification accuracy is 90.1% on average.

The classification accuracy and number of final features reported by our greedy solution are illustrated in Figure 6. For the greedy approach, the number of selected features ranges from 5 to 47 depending on the design parameter λ_D. The optimal feature set has a length of 24.2 features on average. The accuracy ranges from 79.5% for the lowest threshold to 96.7% for λ=0.15, with an average accuracy of 90.0%. We note that, unlike the ILP solution, the greedy approach does not result in a monotonically increasing accuracy curve. The accuracy curve for the greedy approach has a local minimum at $\lambda_D = 0.13$. This is in fact due to the sub-optimality of the greedy approach, which does not necessarily select the optimal feature set at each step.

6.4 Energy Analysis

Figure 7 shows the total energy consumption of the selected features obtained by the ILP and greedy solutions. For optimal case, the energy values range from 33μJ for λ_D = 0.05 to 1184μJ for λ_D = 0.15, resulting in an energy consumption of 467μJ on average. The energy consumption results for the greedy approach range from 33μJ to 1204μJ with an average of 484μJ.

Table 3 lists energy savings obtained by applying our power-aware feature selection technique. For this specific table, only ILP results are reported. For greedy approach, however, similar results are achieved. The energy savings for the greedy solution range from 29.6% for the highest accuracy (96.7%) to 98.1% for the lowest performance (79.5%). This results in an average energy saving of 71.6% using the greedy solution.

Table 3: Energy savings using the proposed method

λ_D	**Saving (%)**	Accuracy (%)
0.05	98.1	80.8
0.07	96.2	85.6
0.09	90.3	89.1
0.11	77.2	92.9
0.13	43.6	95.7
0.15	30.7	96.7
Avg.	72.7	90.1

7. CONCLUSION AND FUTURE WORK

The accuracy and power trade-offs in wearable monitoring systems have been investigated, that is, guaranteed classification accuracy is required, while minimizing the system's power consumption. Our study in this paper accounts for energy consumption in the process of feature selection. To achieve that, we first formulated the problem as a weighted minimum set cover approximation, which is one of the oldest and most studied NP-hard problems. We then devised a greedy approach to select the features needed for the identification of activities being performed in a power-efficient manner. Experimental results on inertial data collected from real subjects demonstrated significant power savings. In the future, we will investigate dynamic feature selection and node activation based on the contextual information about the subject in real-time. That is, information such as subject's location and current activity can be used in real-time to further eliminate power-hungry features and deactivate sensor nodes not contributing to the classification.

8. REFERENCES

[1] P. Lukowicz, O. Amft, D. Roggen, and J. Cheng, "On-body sensing: From gesture-based input to activity-driven interaction," *Computer*, vol. 43, no. 10, pp. 92 –96, oct. 2010.

[2] U. Anliker and et. al., "Amon: a wearable multiparameter medical monitoring and alert system," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 8, no. 4, pp. 415 –427, dec. 2004.

[3] M. Dash and H. Liu, "Feature selection for classification," *Intelligent data analysis*, vol. 1, no. 3, pp. 131–156, 1997.

[4] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," in *Proceedings of the Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11)*. Corvallis, Oregon: AUAI Press, 2011, pp. 266–273.

[5] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, December 2004.

[6] B. French, D. Siewiorek, A. Smailagic, and M. Deisher, "Selective sampling strategies to conserve power in context aware devices," in *Wearable Computers, 2007 11th IEEE International Symposium on*, oct. 2007, pp. 77 –80.

[7] H. Ghasemzadeh, V. Loseu, and R. Jafari, "Collaborative signal processing for action recognition in body sensor networks: a distributed classification algorithm using motion transcripts," in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, ser. IPSN '10. New York, NY, USA: ACM, 2010, pp. 244–255.

[8] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Troster, "Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection," *Lecture Notes in Computer Science*, vol. 4913, p. 17, 2008.