Minimum-Cost Channel Selection in Wearables

Ramesh Kumar Sah
Washington State University

Nooshin Taheri Chatrudi
Arizona State University

Stephanie Marita Carpenter
Arizona State University

Hassan Ghasemzadeh
Arizona State University

***Abstract*—Sensor channel selection is an important optimization problem in resource-constrained wearable systems with the goal of identifying an optimal set of input sensors for efficient machine learning. We introduce a framework for this optimization problem, mathematically formulate the minimum-cost channel selection (MCCS), and propose two novel algorithms to solve the problem. Branch and bound channel selection finds a globally optimal channel subset and the greedy channel selection finds the best intermediate subset based on our proposed penalty function. These proposed channel selection algorithms are conditioned with both performance and the cost of the channel subset. We evaluate both algorithms on two publicly available time series datasets for activity recognition and mental task classification. Branch and bound channel selection achieve a cost saving between 92.6% and 95.7%, and the greedy approach reduces the cost between 51.8% and $91.4,\%$ for performance thresholds of 50% and 70%.**

***Index Terms*—channel selection, sensor systems, time series, machine learning**

I. Introduction

Wearable sensor systems operate in a resource-restricted setting with limited computational power and memory. Machine learning algorithms and wearables are being increasingly utilized in digital health applications. However, oftentimes, different sensing modalities are suitable for achieving a particular goal. For example, consider stress detection for which bio-markers such as heart rate variability, skin conductance response, and core body temperature are suitable. Another example is human activity recognition, in which sensor nodes can be placed at different body sites. In such cases, it becomes important to determine the optimal set of sensor channels to meet the requirements of the machine learning task while adhering to the design and operating limitations of the system.

Sensor channel selection is defined as the identification and removal of channels that provide a negative or negligible contribution to a goal task T. The problem is selecting k channels out of n given channels while optimizing the performance and total cost. Given this problem setting, there exists $C(n,k) = \frac{n!}{k! \times (n-k)!}$ channel subsets. For small search spaces, an exhaustive search can be used to identify and remove redundant sensor channels. However, the search space grows exponentially with the size of channels set (n), and exhaustive search is not feasible for large search spaces. A channel selection algorithm combines a search technique to find new channel subsets and an evaluation method to assess the performance of the selected subset. A commonly used evaluation process involves training a machine learning algorithm for the considered task on the selected subset. The performance of the trained model is used as the proxy score of the selected channel subset. This approach of evaluation is similar to wrapper-based feature selection, which has been extensively studied in machine learning literature. Most prior work on channel selection follows the wrapper-based evaluation paradigm with some heuristic to either limit the search space of channel subsets [ğrul2017determining], [1] or modify the learning process to encourage the model to learn features from least number of input channels during the training process [2], [3]. In general, these methods only consider the performance criteria in their evaluation step, and the cost of the channel subset is not used in the decision-making process. This leaves a gap in the literature regarding the optimal channel subset, which not only meets the performance criteria but also considers the minimum total cost. In this work, we present two novel backward search algorithms to address this knowledge gap by finding a channel subset with minimum cost while ensuring a lower bound on the performance is met.

II. Minimum Cost Channel Selection

Minimum cost channel selection (MCCS) is defined as identifying and selecting a subset of n channels with minimum total cost W while ensuring that the selected channels achieve a lower bound (λ) on the performance according to the given performance function $f(.)$. Furthermore, depending on the learning task T, the performance function f can either be maximized (accuracy) or minimized (mean squared error).

A. Problem Definition

Given n sensor channels $C_n = \{c_1, c_2, \dots c_n\}$ and cost $W_n = \{w_1, w_2, \dots, w_n\}$ for selecting each sensor channel. The MCCS problem is to minimize the total cost

$$\text{Minimize} \sum_{i=1}^{n} w_i a_i \tag{1}$$

Subject to:

$$\begin{aligned} & f(c_1 a_1, c_2 a_2, \dots, c_n a_n) \geq \lambda \\ & a_i \in \{0, 1\} \end{aligned} \tag{2}$$

Here, f is a performance function, a_i is a binary value indicating selection of channel c_i, λ is the lower bound on the performance, and w_i is the normalized cost of selecting channel c_i. Normalized cost is obtained for all channels given W_n such that $\sum_{i=1}^{n} w_i = 1$.

B. Branch and Bound Channel Selection

Let $(c_1, \dots, c_{\bar{s}})$ be the $\bar{s} = n - s$ channels to be discarded to obtain the channel subset C_s of size s. Each channel c_i can take on value in $\{1, 2, \dots, n\}$. Here, the order of c_i's is not important, and we only consider sequences of c_i's such that $c_1 < c_2 < \cdots < c_{\bar{s}}$. The performance function (f) is a function of the selected channel subset (C_s) obtained by discarding $c_1, \ldots, c_{\bar{s}}$ channels from the n channel set. Now, the channel subset selection problem is to find the subset $c_1^*, \ldots, c_{\bar{s}}^*$ to discard such that

$$\begin{aligned} f(C_{\bar{s}}^*) &= \max\ f(C_s^*) \\ \text{and } W(C_{\bar{s}}^*) &= \min\ W(C_s^*) \end{aligned} \tag{3}$$

W is a cost function defined as the sum of the normalized cost of all channels in the selected subset C_s. Let us assume the performance function f satisfies monotonicity defined by

$$f_n(c_1, c_2, \ldots, c_n) \geq f_{n-1}(c_1, c_2, \ldots, c_{n-1}) \geq \cdots \geq f_1(c_1) \tag{4}$$

The monotonicity principle means that a subset of channels should not be better than any larger set containing the subset. We acknowledge that not all types of neural networks satisfy the monotonicity principle, but recent works have shown ways to create deep neural networks with monotonic properties [4]. The cost function already satisfies the principle of monotonicity, i.e., $W_n(c_1, c_2, \ldots, c_n) \geq W_{n-1}(c_1, c_2, \ldots, c_{n-1}) \geq \cdots \geq W_1(c_1)$. Then, given the lower bound (λ) on the value of the performance, we can write

$$\lambda \leq f(C_s^*) \tag{5}$$

And, if $f(C_k)(k > s)$ is less than λ, then from equation 4,

$$\begin{aligned} & f(C_s) \leq \lambda \\ & \forall \quad \{C_{k+1}, \ldots, C_s\} \end{aligned} \tag{6}$$

Equation 6 means that whenever the performance function evaluated for any subset is less than λ, all subsets that are successors of that subset also have performance value less than λ, and therefore cannot be the optimal solution. This forms the basis for the branch and bound channel selection algorithm. The branch and bound method successively generates portions of the solution tree and computes the performance value. Whenever a sub-optimal partial subset satisfies condition 6, the sub-tree under that subset is implicitly rejected, and enumeration begins on the subsets that have not yet been evaluated [5]. Algorithm 1 describes the proposed branch and bound channel selection.

Algorithm 1 Branch and bound channel subset selection

Input: List of channels $C_n = \{c_1, c_2, \ldots, c_n\}$, Cost of each channel $W_n = \{w_1, w_2, \ldots, w_n\}$, and Number of channels n
Parameter: Objective function f and Performance threshold λ
Output: Globally optimal channel subset C^*, cost of the selected channel subset W^*, and list of optimal subsets C_o

Set $C^* = C_n$ and $W^* = \sum w_i$
Create stack S and hash table H
Set current subset node $K_{current}$ = C_n
C_o = []
POPPED $= 1$
if $f(K_{current}) < \lambda$ **then**
    return C^*, W^*, C_o
end if
Push $K_{current}$ into S
Map $Kcurrent$ into H
while S is not empty **do**
    $K_{previous} = K_{current}$
    Create children subset nodes of $K_{current}$ and store them in the ascending order of the cost in L
    for subset n in L **do**
        $K_{current} = n$
        Check the performance of $K_{current}$ and update S, H
        Update C^*, W^* and C_0 if needed
    end for
    if $K_{current} == K_{previous}$ **then**
        Pop S and assign to $K_{current}$
        POPPED $= 1$
    else
        POPPED $= 0$
    end if
end while
return C^*, W^*, C_o

C. Greedy Channel Selection

The branch and bound algorithm assume monotonicity in performance which may not be always true. Furthermore, in the worst case branch and bound search must evaluate all possible channel subsets and consequently will have an exponential runtime [6]. In light of these limitations, we also propose a greedy algorithm 2 for sub-optimal channel subset selection.

Let C be a root channel subset node and $C - c$ be its children subset node. The subset $C - c$ is created by discarding the channel c from the parent subset C. We define a *Penalty* function

$$penalty = \alpha \times f(C - c) + (1 - \alpha) \times W(C - c) \tag{7}$$

where $f(C - c)$ is the value of the performance function on the channel subset $C - c$ and $W(C - c)$ is the sum of the normalized cost of channels in the subset $C - c$. α is a balancing term used to control the influence of performance and cost on the penalty value. Given the penalty function, the greedy algorithm selects a channel subset that meets the performance threshold i.e., $f(C - c) \geq \lambda$ and has the minimum value of penalty at each intermediate stage. Also, since the goal is to minimize the value of the penalty function, for classification problems we modify the penalty function as

$$penalty = \alpha \times (1 - f(C - c)) + (1 - \alpha) \times W(C - c) \tag{8}$$

The algorithm greedily selects a subset with the least penalty value and hence is able to achieve a runtime of $O(n^2)$.

D. Cost Model

Cost model defines the cost of a channel based on the some input parameters such as: 1) computation and memory requirement which are directly related to sampling frequency, 2) power requirement, 3) sensing requirement, 3) usability and

Algorithm 2 Greedy channel subset selection

Input: List of channels $C_n = \{c_1, c_2, \ldots, c_n\}$, Normalized cost of each channel $W_n = \{w_1, w_2, \ldots, w_n\}$
Parameter: Objective function f, Performance threshold λ, and α
Output: Locally optimal channel subset C and cost W

Set $C = C_n$ and $W = \sum w_i$
Set current subset node $C_k = C$
if $f(C_k) < \lambda$ **then**
    return C, W
end if
Set best penalty $S_b = inf$
while true **do**
    $L =$ children subset nodes of C_k
    if L is Empty **then**
        return C, W
    end if
    for subset S in L **do**
        Check performance of S
        Update C and W if needed
    end for
end while
return C, W

interpretability cost, 4) manufacturing cost, and 5) other cost. In our analysis, we generate the cost for each channel using a simple heuristic based on the sampling frequency of the sensor channel. Sensor channel with higher sampling frequency are assigned a larger cost and vice-versa. In practice, the cost of the sensor channel can be determined as needed and used with our proposed algorithms to determine the optimal channel subset.

III. Analysis and Results

A. Datasets

EEG Mental Task [7] dataset contains electroencephalogram (EEG) signals recorded for binary mental arithmetic task detection using Neurocom EEG 23-channel system at a sampling frequency of 500 Hz. A high-pass filter with a 30 Hz cut-off frequency and a power line notch filter (50 Hz) were used to eliminate noise and artifacts from all EEG channels. All recordings are artifact-free segments of 60 seconds in duration. We further subdivided the segments into input windows of size 10 seconds with 5 seconds overlap between consecutive windows.

PAMAP2 [8] is a human activity recognition dataset recorded from 9 participants wearing 3 sensor units (chest, wrist, and ankle) and performing 18 activities. Each sensor unit contained $3-$axis accelerometer, gyroscope, and magnetometer all sampled at 100 Hz. Altogether there are 27 sensor channels with 9 channels from each body location. The signal from each sensor is subdivided into windows of size 30 seconds with 15 seconds overlap between consecutive windows. The activity recognition task is defined as a 7 class classification problem.

B. Model Architecture

We have used $1D$ Convolutional Neural Network (CNN) architecture to evaluate each channel subset during the search process, as shown in Fig 1. CNNs are known to work well for time-series classification problems [9] and can be trained with raw sensor values without feature computation and selection. The modular architecture of CNN can also accommodate dynamic changes in input channels. Each input channel in the considered subset is assigned a separate feature extraction block, and outputs from all feature extractors are aggregated in the classification block to learn the mapping from input to output.

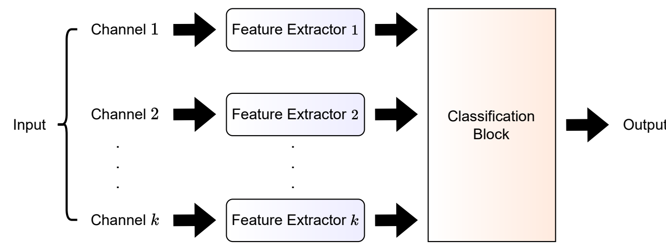

Fig. 1: Modular architecture of the CNN model with a number of feature extractors equal to the number of input channels.

The feature extraction block consists of two one-dimensional convolutional layers, and the classification block has two fully-connected layers. ReLU activation is used in all intermediate layers, and Softmax activation in the output layer consists of the same number of neurons as the number of output classes. In all cases, the model is trained for 100 epochs using cross-entropy loss and Adam optimizer with a learning rate set to 0.001.

C. Channel Subset Selection

In our analysis, each axis of a sensing modality is considered an independent sensor and we recommend consulting original publications for more details about sensor channels in both datasets [7], [8]. We initially set $\alpha = 0.5$ for greedy channel search and measured the performance in terms of accuracy of the trained model. Table I shows the optimal channel subset for the EEG mental task dataset determined using branch and bound, and greedy channel selections. Since the sampling frequency of all channels in the EEG dataset is equal, the normalized cost of each channel is also equal and set to $w_i = 0.043$. The accuracy of the model trained on all available channels is considered baseline performance and was 73.48%. Given the baseline performance, the performance threshold λ was set to 0.7 or 70% accuracy. For the EEG dataset, branch and bound channel selection was able to achieve a cost saving of 95.7%, and the greedy search was able to reduce the cost by 91.4%.

TABLE I: Optimal channel subset for EEG dataset determined using branch and bound (B&B) and greedy methods.

Method	Selected Subset	Accuracy (%)	Cost	Penalty	Cost Savings
B&B	FP1	70.31	0.043	0.169	95.7%
Greedy	(C3, F3)	72.33	0.086	0.191	91.4%

All channels in the PAMAP2 dataset also have equal sampling frequency, and consequently, the cost of each channel is equal and set to $w_i = 0.037$. The baseline performance accuracy was determined to be 59.02%, and the performance threshold λ was set to 0.5 or 50% accuracy. Branch and bound search selected the subset with 2 channels with performance 51.22%, cost 0.074, and penalty 0.282 to be globally optimal. A subset with 13 channels was determined to be best with greedy search. The cost of the greedy subset is 0.4814 with a performance of 89.75% and a penalty of 0.2919. A cost saving of 92.6% was achieved with branch and bound search and a cost saving of 51.9% with greedy search.

D. Effects of Alpha

We set $\alpha = 0.5$ in the preceding analysis, placing equal importance on the cost and performance for greedy channel selection. However, in practice, minimizing cost might be more important than maximizing performance and vice-versa. Fig 2 shows the performance and cost of the selected channel subset at different values of α. Larger values of α put greater emphasis on the performance and smaller values of α favor channel subsets with lower costs. For both datasets, at larger values of α, the accuracy of the selected subset is higher, but the cost is also high. This is expected because a greater number of input channels will provide more information to the model to learn the mappings between input and outputs, consequently increasing the performance.

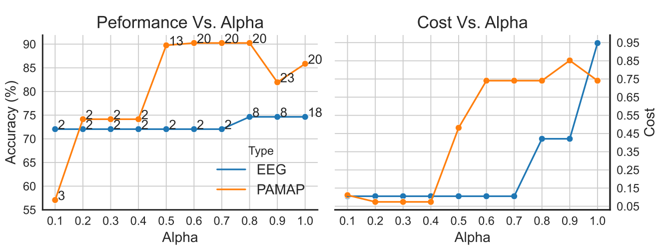

Fig. 2: Accuracy and cost of the selected channel subset using greedy search for both datasets at different values of alpha. The values on the line denote the number of selected channels.

E. Comparison Analysis

Channel selection approaches either use exhaustive search [ğrul2017determining], [1] to evaluate all possible combinations of channels or some optimization method [3], [10] to limit the search space and select best-performing channel subsets. A direct comparison of our approach with existing studies is not feasible because existing studies often lack a provision for performance threshold and do not consider the cost of channels in their method. However, to facilitate comparative analysis, we have compared the results reported in [10] for the PAMAP2 dataset against results obtained using our algorithms. Authors in [10] devised a deep neural network to promote learning from the lowest number of channels and showed superior performance compared to other channel selection approaches [2], [3]. The algorithm in [10] selected 15-channels for f1-score of 0.88. Our greedy algorithm selects 13-channels for the accuracy of 89.75% and the branch and bound approach selected 2-channels to meet the performance threshold of 50% with minimum cost. Given that the cost assigned to each channel is equal in our evaluation, the comparison is valid and is not affected by the cost parameter.

IV. Conclusion

In this work, we have proposed and validated two sensor channel selection algorithms to determine an optimal subset of channels that meets the performance criteria with minimum cost. Proposed algorithms can be used in real-life applications to optimize the cost of a sensor system while also ensuring a performance guarantee. Branch and bound channel selection also allows for dynamic selection of channels during run-time since it returns a list of channel subsets satisfying the performance threshold. When some channels from the globally optimal subset become unavailable during run-time, channels from the next best subsets can be used to keep the system operational. Also, our evaluation scheme is model agnostic, and any other type of learning algorithm can be used instead of CNN.

Acknowledgment

This work was supported in part by grant IIS-1954372 from the National Science Foundation.

References

[1] M. Espinilla, J. Medina, A. Calzada, J. Liu, L. Martínez, and C. Nugent, "Optimizing the configuration of an heterogeneous architecture of sensors for activity recognition, using the extended belief rule-based inference methodology," *Microprocessors and Microsystems*, vol. 52, pp. 381–390, 2017.

[2] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster, "Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection," in *Wireless Sensor Networks: 5th European Conference, EWSN 2008, Bologna, Italy, January 30-February 1, 2008. Proceedings*. Springer, 2008, pp. 17–33.

[3] J. Cao, W. Li, C. Ma, and Z. Tao, "Optimizing multi-sensor deployment via ensemble pruning for wearable activity recognition," *Information Fusion*, vol. 41, pp. 68–79, 2018.

[4] D. Runje and S. M. Shankaranarayana, "Constrained monotonic neural networks," *arXiv preprint arXiv:2205.11775*, 2022.

[5] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Transactions on Computers*, vol. 26, no. 09, pp. 917–922, 1977.

[6] W. Zhang, *Branch-and-bound search algorithms and their computational complexity*. University of Southern California, Information Sciences Institute, 1996.

[7] I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, and O. Shpenkov, "Electroencephalograms during mental arithmetic task performance," *Data*, vol. 4, no. 1, p. 14, 2019.

[8] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *2012 16th International Symposium on Wearable Computers*. IEEE, 2012, pp. 108–109.

[9] R. K. Sah, M. J. Cleveland, A. Habibi, and H. Ghasemzadeh, "Stressalyzer: Convolutional neural network framework for personalized stress classification," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 4658–4663.

[10] C. F. S. Leite and Y. Xiao, "Optimal sensor channel selection for resource-efficient deep activity recognition," in *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021)*, 2021, pp. 371–383.