Packet Pruning: Finding Better Energy Spending Policies for Batteryless Human Activity Recognition

Geffen Cooper
The University of Texas at Austin
geffen@utexas.edu

Radu Marculescu
The University of Texas at Austin
radum@utexas.edu

2024 IEEE 20th International Conference on Body Sensor Networks (BSN) | 979-8-3315-3014-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/BSN63547.2024.10780463

***Abstract*—Batteryless sensing devices which rely on energy harvesting can enable more sustainable and long-lasting Internet of Things (IoT) based wearables. While it has become feasible to implement energy-harvesting based wearables for digital health applications, it remains challenging to integrate such devices and the data they collect into machine learning pipelines for tasks such as human activity recognition (HAR). A key obstacle is uncertainty in the data acquisition process. Given the discontinuous and uncertain availability of harvested energy, *when* should a sensor spend energy to sample and transmit data packets for processing? A common approach is to spend energy *opportunistically* by sending packets whenever sufficient energy is available. However, when considering a specific task, namely HAR with kinetic energy harvesting based sensors, this approach unfairly prioritizes data from activities where more energy can be harvested (e.g., running). In this work, we improve the opportunistic energy spending policy by *pruning* redundant packets to reallocate energy towards activities where less energy is harvested. Our approach results in an increase in the F1-score of 'lower energy' activities while having a minimal impact on the F1-score of 'higher energy' activities.[1]**

***Index Terms*—Batteryless sensing, deep learning, HAR**

I. Introduction

Batteryless sensing has emerged as a popular alternative to battery-powered Internet of Things (IoT) devices in an effort to increase sustainability and ensure long-lasting, maintenance-free operation [1]. Batteryless IoT devices range from complex systems with local processing [2] to simple beacons [3] targeted towards wearable health applications. This work targets the latter category which views batteryless sensors as *optimizable* data sources in a broader edge AI system for deep learning-based analytics. Specifically, we focus on human activity recognition (HAR) using data from wearable kinetic energy harvesting (KEH) based sensors. Batteryless HAR seeks to enable long-term activity monitoring by providing an autonomous, wearable data source which does not require charging or replacement. Such characteristics are important for wearables in health monitoring [4] and recent work has demonstrated promising results of using long-term data from wearable activity trackers to detect mental health disorders [5].

With advances in energy harvesting technologies, multiple KEH-based HAR devices have been proposed in the past several years [3], [6], [7], [8]. These works demonstrate the feasibility of building a wearable, batteryless KEH device for HAR. However, they have several significant limitations that make them unsuitable for complex, continuous activity recognition settings. First, these works do not use an accelerometer as a data source, and instead leverage signals derived from the KEH device such as the time between packet transmissions. These signals are not informative over short time periods (less than one second) and require relatively long windows (e.g. tens of seconds) in order to be used for activity classification. Second, these works leverage traditional classifiers like decision trees which are not well suited for sparse, asynchronous, and variable length packet sequences. Finally, these works use *opportunistic* energy spending policies which sample data whenever sufficient energy is available, not necessarily when the data is 'informative' for activity recognition. Therefore, existing approaches are limited to activity sequences where 'high-motion' activities occur in long isolated segments.

In this work, we consider the more challenging setting of *dense activity prediction* with a diverse sequence of high and low-motion activities which may occur with variable duration. The KEH device is treated as a Bluetooth (BLE) beacon which samples and transmits short accelerometer packets (8 samples at 25 Hz). Given these challenges, and the limitations of existing approaches, we build on our preliminary work [9] and leverage a more advanced activity classifier based on modern convolutional neural networks (CNN) and transformers which can handle sparse, asychronous data. Given a pretrained activity classifier, **the central challenge we address in this work** is how to *learn* an energy spending policy which determines *when to send* packets based on data redundancy. This is a significant departure from prior work on energy allocation for KEH-based HAR [10] which considers a much simpler setting in which data can constantly be sampled (not batteryless) and energy consumption is controlled by altering the sampling rate.

II. Methods

A. Problem Setup

Our problem setting is derived from [9]. The data acquisition and HAR classification pipeline is shown in Fig. 1. Multiple KEH-based wearable sensors are worn on the body and accumulate energy from human motion. When sufficient energy has been harvested at a given sensor, an energy spending policy determines whether to sample and transmit a packet of accelerometer data to a local edge device such as a mobile phone which uses deep learning for HAR.

[1]Code: `https://github.com/SLDGroup/PacketPruning`

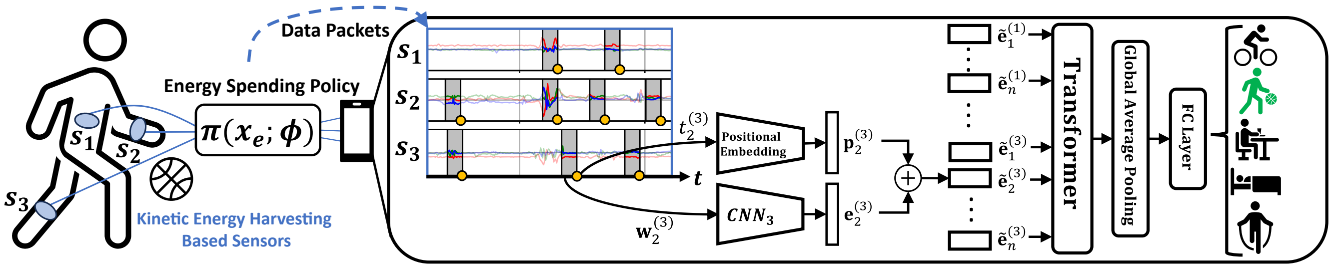

Fig. 1. Batteryless HAR Pipeline. Data from multiple KEH-based sensors gets sampled and transmitted to a nearby edge device based on an energy spending policy. The sparse, asychronous data from these sensors gets classified by a transformer based architecture which accounts for spatial and temporal context.

Energy Harvesting and Consumption Model: We estimate the kinetic energy that can be harvested from accelerometer traces using a mass-spring-damper model as described in [11]. For energy consumption, we follow the model used in [9]. From an off state, the device must harvest E_{init} to turn on and initialize. To acquire data (sampling and BLE transmission), the device must harvest E_{packet}. At all other times the device is in a low-power state and has idle power consumption P_{idle}.

Deep Learning Architecture: Given that packets from each sensor arrive in an asychronous manner, standard HAR approaches [12], [13] which assume time synchronized data cannot be used. Thus, we leverage the advanced deep learning architecture shown in Fig. 1. For each body part, only the n most recent packets are buffered and packets older than T seconds are removed from the buffer. Each packet is passed through a CNN to extract a feature embedding. These features are augmented with a positional embedding which learns spatial and temporal context based on which body part they arrived from and their arrival time relative to the oldest packet. The packet embeddings are processed using a transformer [14] and aggregated using global average pooling; this is followed by a fully connected layer for classification. We train the model by sampling segments from the training activity sequence. These segments are 'sparsified' into packets by applying an opportunistic policy with random delays to ensure that a diverse set of packets are observed.

B. Packet Pruning Methodology

A limitation of an opportunistic energy spending policy is that it implicitly prioritizes data from activities where more energy gets harvested. To understand this effect on HAR classification, we plot the F1-score of each activity as a function of the fraction of the total packets sampled during each activity in the DSADS [15] dataset. In Fig. 2 we observe a weak positive correlation, reflecting that low-motion activities like sitting get lower F1-scores while high-motion activities like jumping get higher F1-scores. However, some activities achieve high F1-scores with low sampling priority and low F1-scores with high sampling priority; this suggests that a more complex energy spending policy is required to sample activities at more informative times. Thus, we propose to improve the opportunistic policy by *pruning* redundant packets in an effort to reallocate energy usage from high-motion activities to other activities.

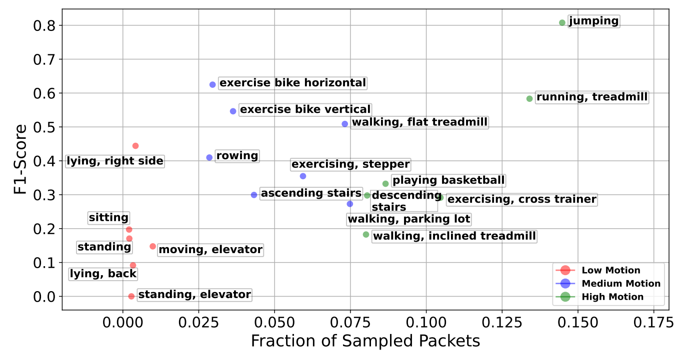

Fig. 2. F1-score of each activity as a function of the fraction of packets sampled in the DSADS dataset. More energy is harvested from high-motion activities so a larger fraction of the total packets are sampled. Red dots represent low-motion activities. Blue dots resemble average-motion activities. Green dots represent high-motion activities.

Offline Approach: We first develop an offline algorithm for packet pruning which assumes access to present and past data. Consider Fig. 3 which shows a window of packets arriving from the left arm and the right leg for a segment of the walking activity when applying the opportunistic policy. The red $\times$ shows a packet to prune which results in a further accumulation of energy, as shown by the red portion of the black line. Given a set of packets, how can we determine which one to prune? Ideally, we want to prune those packets which will decrease the average classification loss (e.g., cross entropy) the most over the sequence of predictions within a segment. We take inspiration from work in the natural language processing community [16] which seeks to determine how flipping words in an input sequence affects the loss for language models. Our problem setting is analogous in that we are flipping packets in a binary manner.

More formally, consider the feature embeddings from a sequence of packets as $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \cdots \mathbf{e}_n]$ where $\mathbf{e}_i \in \mathbb{R}^d$ and d is the embedding dimension. When pruning packets, we are applying a mask $\mathbf{m} = [m_1, m_2, \cdots m_n]$ where $m_i \in \{0, 1\}$ such that the new packet sequence becomes $[\mathbf{e}_1, \mathbf{e}_2, \cdots \mathbf{e}_n] \odot [m_1, m_2, \cdots m_n]$. Before pruning, the mask is $\mathbf{m}^{(o)} = [1, 1, \cdots 1]$ since all packets are present; after pruning, the mask has $\mathbf{m}^{(p_i)} = [1, \cdots, 1, 0, 1, \cdots, 1]$, where a zero

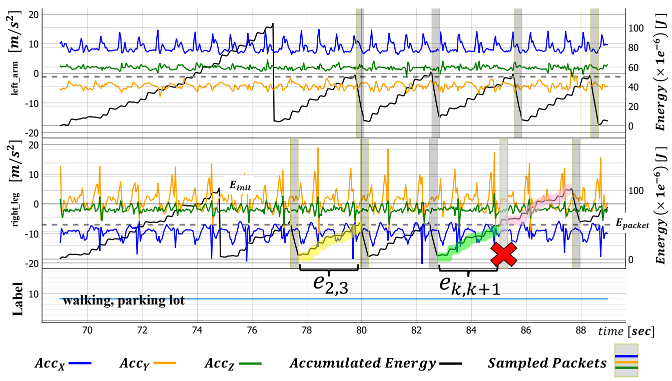

Fig. 3. Segment of simulated data acquired from two body parts. The blue, yellow, and green lines show the accelerometer X, Y, and Z signals respectively over time. The black line shows the accumulated energy over time on a second Y-axis. The dashed gray line shows the E_{packet} energy threshold for sending a packet. The shaded gray regions show packets that are sampled while the red $\times$ shows a packet that gets pruned. The yellow and green highlighted regions show two examples of energy traces between successive time steps in which the transmit threshold is reached.

means the packet in the i^{th} entry is pruned. Our goal is to find the mask $\mathbf{m}^{(p_i)}$ such that when it gets applied to the packet sequence $\mathbf{E}$, the average loss over the activity sequence corresponding to $\mathbf{E}$ decreases the most. We express this as:

$$L = \sum_{t=t_s}^{t_e} CE\Big(y_t, f(\mathbf{E}_{r_t} \odot \mathbf{m}_{r_t})\Big) \tag{1}$$

where t_s, t_e represent the start and end of an activity segment we want to prune packets from, CE is the cross entropy loss, y_t is the label at time t, f is the classifier, and $\mathbf{E}_{r_t}, \mathbf{m}_{r_t}$ represent a subset of the packet and mask sequence which are active for the prediction at time t. Recall, that only the n most recent packets since time t are considered as input to the model; here we denote this using the subscript r_t.

When pruning a packet, we move the mask from $\mathbf{m}^{(o)}$ to $\mathbf{m}^{(p_i)}$, which pushes the mask in the direction $\mathbf{m}^{(op_i)} = \mathbf{m}^{(p_i)} - \mathbf{m}^{(o)} = [0, \cdots, 0, -1, 0, \cdots, 0]$. To estimate the impact on the loss, we evaluate the directional derivative along this vector as $\nabla_{\mathbf{m}^{(op_i)}} L = \nabla_{\mathbf{m}^{(o)}} L \cdot (\mathbf{m}^{(p_i)} - \mathbf{m}^{(o)}) = -\frac{\partial L}{\partial \mathbf{m}_i^{(o)}}$ which is the negative of the i_{th} component of the gradient of the loss with respect to $\mathbf{m}^{(o)}$. Since we want to find $\mathbf{m}^{(p_i)}$ which *decreases* the loss the most, we consider the negative gradient which cancels the negative from $-\frac{\partial L}{\partial \mathbf{m}_i^{(o)}}$. Overall, we prune the packet based on the largest entry of the gradient of the loss with respect to $\mathbf{m}^{(o)}$: $\underset{i}{\text{argmax}}\ \nabla_{\mathbf{m}^{(o)}} \Big(\sum_{t=t_s}^{t_e} CE\Big(y_t, f(\mathbf{E}_{r_t} \odot \mathbf{m}_{r_t})\Big)\Big)$.

Online Approach: In practice, we cannot use the offline approach as there is no benefit to pruning a packet after it is sampled (since the energy has already been consumed). Regardless, we find the offline approach to be useful in understanding which types of packets are less informative. For example, in Fig. 3, the offline approach determines that the fourth packet should be pruned which does indeed result in a minimal loss increase. From Fig. 3 it can be seen that this packet is redundant; the classifier has sufficient information from the first three packets to predict the activity. Given this insight, we propose a heuristic for detecting redundant packets based of the energy traces between successive transmissions.

Formally, consider a sequence of the k most recent packets as $[p_1, p_2, \cdots, p_k]$ and define the energy traces between these packets as $[e_{1,2}, e_{2,3}, \cdots, e_{k-1,k}]$ (see Fig. 3). Upon reaching a sufficient amount of energy for packet $k+1$, we determine whether to prune this packet (i.e., do not sample it) by comparing the similarity of the energy trace $e_{k,k+1}$ with the k most recent ones. Given that these traces can be of different lengths, we extract a simple feature vector $[mean, variance, skew, kurtosis]$ and compute the cosine similarity between the most recent trace and the average of the k most recent traces. If the similarity exceeds a threshold, we determine that the packet is redundant and do not sample it. As a result, the energy accumulates over time. In this way, each packet pruned reallocates the energy to the next activity in the sequence which may be a low-motion activity.

C. Experimental Details

For all experiments we consider a pretrained transformer model as described in section II-A. The goal is to improve the model's predictions by providing it more informative data through better energy spending policies. The model is trained on data from an opportunistic policy with random delays added to each packet. This ensures that a diverse combination of packet sequences are observed during training.

Dataset Preprocessing: We train and evaluate our model on the Opportunity [17] (5 classes, 7 body parts) and DSADS [15] (19 classes, 5 body parts) datasets. The Opportunity data is collected from natural activity sequences. For DSADS, we create synthetic sequences by sampling randomly sized segments (10-30 sec each) from each activity. For the Opportunity dataset, we use data from run 2 from user 1 for validation, runs 4 and 5 from users 2 and 3 for testing, and all remaining runs for training as done in prior HAR works [13]. Similarly, for DSADS we use data from user 6 for validation, users 7 and 8 for testing, and all other users for training.

Evaluation: We evaluate each approach using dense activity prediction which requires the model to make a prediction at every time step regardless of whether new data packets have arrived. Given the label and prediction sequence, we calculate and report the macro-averaged F1-score.

III. Results and Discussion

We compare three energy spending policies in Table I. The opportunistic policy samples a packet whenever sufficient energy is available. The random policy adds random delays to each opportunistic packet. The online pruning approach uses the energy trace to determine if a packet is redundant and should be pruned. When a packet is pruned, the energy level accumulates into a surplus state which can reduce the delay between the start of an activity and the arrival of the first packet. Otherwise, sufficient energy must be accumulated after

TABLE I
MACRO F1-SCORE FOR DENSE PREDICTION (HIGHER IS BETTER)

Method	Opportunity	DSADS
Opportunistic Policy	53.09	34.54
Random Policy	55.69	34.55
Online Pruning	55.56	39.28

activity transitions which can result in segments of missing data. Also, each time the device incurs a power failure, there will be a delay until the device harvests E_{init} to turn on.

As shown in Table I, the pruning based approach outperforms the opportunistic policy. To understand the fine-grained benefits of the pruning based approaches, in Fig. 4 we visualize the F1-score for each individual activity as a function of the fraction of total packets sent for the DSADS dataset. From Fig. 4 we can see that pruning has a direct positive impact on most of the low-motion activities. Indeed, by increasing the fraction of packets allocated to these activities, we can significantly increase the F1-score in comparison to the opportunistic policy. We also observe that across all other activities there is little degradation in the F1-score and some activities actually improve even with fewer packets sampled. The improvement for the Opportunity dataset is smaller. While this dataset has fewer activities, they often occur in short durations (3-5 seconds) which makes energy allocation challenging given that energy is being accumulated over successive activities rather than long segments. This may explain why the random policy performs similarly to targeted pruning.

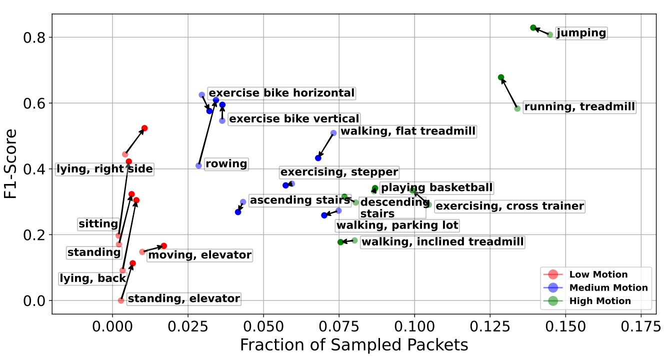

Fig. 4. F1-score of each activity as a function of the fraction of packets sampled in the DSADS dataset. This is an extension of Fig. 2 that shows how each dot moves when changing the energy spending policy from opportunistic (base of the arrow) to the online pruning approach (tip of the arrow).

IV. CONCLUSION

In this paper, we have developed a packet pruning methodology to enhance the performance of batteryless HAR from the perspective of energy spending policies. Our results explicitly show that opportunistic policies do not sufficiently sample data from low-motion activities. Our pruning based approach directly increases the fraction of sampled packets, and thus the F1-score of low-motion activities, without degrading the F1-score on other activities. In future work, we seek to generalize this pruning approach into a full optimization problem for optimal packet sampling. This would enable richer energy spending policies that sample packets at arbitrary time steps, beyond those in which the transmit threshold is hit.

ACKNOWLEDGMENT

This work was supported in part by NSF Grant ECCS-2428656 and the iMAGiNE Consortium at UT Austin.

REFERENCES

[1] S. Ahmed, B. Islam, K. S. Yildirim, M. Zimmerling, P. Pawełczak, M. H. Alizai, B. Lucia, L. Mottola, J. Sorber, and J. Hester, "The internet of batteryless things," *Communications of the ACM*, vol. 67, no. 3, 2024.

[2] A. Bakar, R. Goel, J. de Winkel, J. Huang, S. Ahmed, B. Islam, P. Pawełczak, K. S. Yıldırım, and J. Hester, "Protean: An energy-efficient and heterogeneous platform for adaptive and hardware-accelerated battery-free computing," in *Proc. of the 20th ACM Conference on Embedded Networked Sensor Systems*, 2022, pp. 207–221.

[3] Z. Chen, L. Teng, L. Xu, J. Yu, and J. Liang, "Mp-har: A novel motion-powered real-time human activity recognition system," *IEEE Internet of Things Journal*, 2023.

[4] L. Lu, J. Zhang, Y. Xie, F. Gao, S. Xu, X. Wu, Z. Ye *et al.*, "Wearable health devices in health care: narrative systematic review," *JMIR mHealth and uHealth*, vol. 8, no. 11, p. e18907, 2020.

[5] R. Dai, T. Kannampallil, S. Kim, V. Thornton, L. Bierut, and C. Lu, "Detecting mental disorders with wearables: A large cohort study," in *Proc. of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation*, 2023, pp. 39–51.

[6] P. Mayer, M. Magno, and L. Benini, "Energy-positive activity recognition-from kinetic energy harvesting to smart self-sustainable wearable devices," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 5, pp. 926–937, 2021.

[7] M. Deumer, M. Sandhu, S. Khalifa, B. Kusy, K. Geissdoerfer, M. Zimmerling, and R. Jurdak, "A battery-free wearable system for on-device human activity recognition using kinetic energy harvesting," in *EWSN'23: Proc. of the 2023 International Conference on Embedded Wireless Systems and Networks*, 2023.

[8] S. Khalifa, G. Lan, M. Hassan, A. Seneviratne, and S. K. Das, "Harke: Human activity recognition from kinetic energy harvesting data in wearable devices," *IEEE Transactions on Mobile Computing*, vol. 17, no. 6, pp. 1353–1368, 2017.

[9] G. Cooper, T. Huang, and R. Marculescu, "Demo abstract: A prototype for machine learning with batteryless sensors," in *2024 IEEE/ACM Ninth International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 2024, pp. 223–224.

[10] L. Xiao, Y. Meng, X. Tian, and H. Luo, "Energy allocation for activity recognition in wearable devices with kinetic energy harvesting," *Software: Practice and Experience*, vol. 51, no. 11, 2021.

[11] M. Gorlatova, J. Sarik, G. Grebla, M. Cong, I. Kymissis, and G. Zussman, "Movers and shakers: Kinetic energy harvesting for the internet of things," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 8, pp. 1624–1639, 2015.

[12] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[13] A. Abedin, M. Ehsanpour, Q. Shi, H. Rezatofighi, and D. C. Ranasinghe, "Attend and discriminate: Beyond the state-of-the-art for human activity recognition using wearable sensors," *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, 2021.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[15] B. Barshan and K. Altun, "Daily and Sports Activities," UCI Machine Learning Repository, 2013, DOI: https://doi.org/10.24432/C5C59F.

[16] J. Ebrahimi, A. Rao, D. Lowd, and D. Dou, "Hotflip: White-box adversarial examples for text classification," in *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2018, pp. 31–36.

[17] D. Roggen, A. Calatroni, L.-V. Nguyen-Dinh, R. Chavarriaga, and H. Sagha, "OPPORTUNITY Activity Recognition," UCI Machine Learning Repository, 2012, DOI: https://doi.org/10.24432/C5M027.