Domain-Informed Label Fusion Surpasses LLMs in Free-Living Activity Classification

Shovito Barua Soumma[1], **Abdullah Mamun**[1,2], **Hassan Ghasemzadeh**[1]

College of Health Solutions[1], Arizona State University
School of Computing and Augmented Intelligence[2], Arizona State University
Phoenix, AZ 85054, USA
{shovito, a.mamun, hassan.ghasemzadeh}@asu.edu

Abstract

FuSE-MET[1] addresses critical challenges in deploying human activity recognition (HAR) systems in uncontrolled environments by effectively managing noisy labels, sparse data, and undefined activity vocabularies. By integrating BERT-based word embeddings with domain-specific knowledge (i.e., MET values), FuSE-MET optimizes label merging, reducing label complexity and improving classification accuracy. Our approach outperforms the state-of-the-art techniques, including GPT-4, by balancing semantic meaning and physical intensity.

1 Introduction

Inferring human activities using wearable sensor data has garnered significant research attention, especially in the realm of medical applications. However, existing activity recognition systems are designed for controlled environments, often with healthy participants, leaving a gap in models that can operate effectively in free-living environments, especially for individuals with chronic conditions (Ermes et al. 2008; Fullerton, Heller, and Muñoz-Organero 2017).

Participants of user studies for activity recognition in free-living environments perform their daily activities and submit open-ended text labels. This brings challenges of having noisy labels and different wordings for the same or similar activities, which makes the activity recognition problem more challenging. To overcome this, we introduce ***FuSE-MET*** (**Fu**sion of **S**emantic **E**mbeddings and **MET** Values), a novel framework that optimizes activity recognition in uncontrolled environments. Unlike previous approaches, FuSE-MET leverages domain-specific knowledge, i.e., metabolic equivalent of task (MET), with semantic meanings to reduce the label space for more efficient classification. By applying a lambda-weighted fusion of BERT-based word embeddings (Devlin et al. 2019) and MET values, the framework effectively balances semantic meaning and physical intensity (PI), addressing challenges such as label disparity and sparsity. This method enables automatic detection and merging of similar activities, resulting in reduced label space while maintaining high classification performance and the integrity of the labels.

[1]**Soure code:** https://github.com/shovito66/FUSE-MET

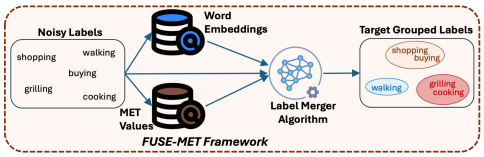

Figure 1: Workflow of FuSE-MET

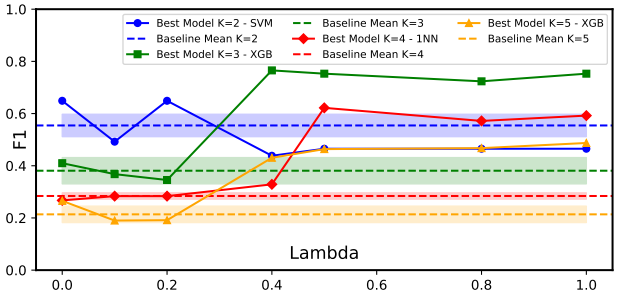

Figure 2: Best models with FuSE-MET vs. the average of models with baseline (no-fusion) for different clusters (K).

2 FuSE-MET Architecture

FuSE-MET restructures the label space generated from user-annotated sensor data collected in free-living environments. Unlike conventional methods that rely solely on semantic similarities, FuSE-MET merges labels by incorporating both semantic meaning and domain knowledge, particularly activity intensity (MET) (Mirzadeh et al. 2019).

We define label merging as an optimization task where we aim to transform noisy labels into meaningful clusters. Let $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$ represent the data, where x_i is the sensor data and y_i denotes user-provided labels from a set $L_{user} = \{a_1, a_2, ..., a_n\}$ of n unique activities. Our goal is to create a reduced label space $L_{merge} = \{l_1, l_2, ..., l_k\}$ with $k \leq n$, where each cluster C_i contains labels that are both semantically and physically similar. Each label a_i is represented by a feature vector $f(a_i) = (1 - \lambda) \cdot w(a_i) + \lambda \cdot m(a_i)$, where $w(a_i) \in \mathbb{R}^d$ is BERT-derived word embedding and $m(a_i)$ is a d dimensional representation of the MET value of the activity. Clustering is performed by minimizing the total within-cluster sum of squared dis-

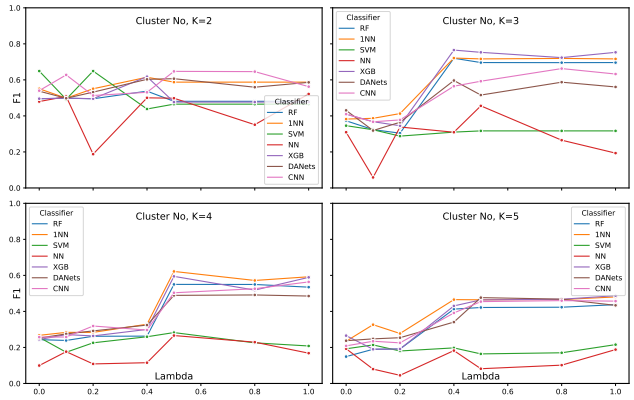

Figure 3: Domain knowledge (λ) vs. F1 for different models.

Algorithm 1: FuSE-MET algorithm. $w_a \in \mathbb{R}^d$, $m_a \in \mathbb{R}^d$, $f_a \in \mathbb{R}^d$. a' is the closest label to a in the MET database. $Decoder : \mathbb{R}_+ \rightarrow \mathbb{R}^d$ creates vectors from met values.

Input: Noisy labels L_{user}, #clusters K, domain coefficient λ, BERT embedding model B, MET values M
Output: clustering labels for L_{merge}
Begin
initialize F as an empty matrix of feature vectors.
for all label a in L_{user} **do**
  $w_a = B(a)$, the word vector of a
  $a' = \text{argmin}_{a''}(cos\_dist(w_a, B(a'')))$ $\forall\, a'' \in M$
  $m_a = Decoder$(normalized MET value of a')
  $f_a = (1-\lambda) \times w_a + m_a \times \lambda$
  add f_a to feature vectors F
end for
return $L_{merge} = K$ clusters by doing k-means on F
End

tances: $C^* = \underset{C}{\text{argmin}} \sum_{i=1}^{k} \sum_{a_j \in C_i} \|f(a_j) - \mu_i\|^2$, where μ_i is the centroid of C_i. This optimization ensures that activities with similar semantic meaning and physical intensity are grouped together, effectively reducing label complexity.

3 Experimental Setup and Analysis

We evaluated FuSE-MET using data from a clinical study involving patients with cardiovascular disease. Accelerometer and gyroscope data were collected from smartphones, and user-provided activities were recorded using active learning. The data were segmented into 5-second windows, with extracted features such as signal intensity, variance, etc. During data collection in uncontrolled environments, several challenges arise, including: (i) Noisy Labels: BERT handles multi-word, inconsistent labels, reducing noise in the label space; (ii) Low Sampling Rate: Despite a sparse sampling interval (5s of data, 5s of rest), FuSE-MET maintains higher performance than GPT-4 by balancing semantic and domain knowledge through MET integration (Table 1). Despite sparse sampling (5 seconds of data followed by 5 seconds of rest), our method maintains high classification accuracy, proving effective even under low-resolution data conditions.

K	Best λ	Best Classifier	ACC	PRE	REC	F1
2	0.2	SVM	**0.98**	**0.69**	**0.62**	**0.65**
	–	*GPT4+CNN	0.62	0.61	0.6	0.59
3	0.4	XGB	**0.68**	**0.77**	**0.76**	**0.77**
	–	*GPT4+XGB	0.63	0.57	0.51	0.51
4	0.5	1NN	**0.69**	**0.63**	**0.62**	**0.62**
	–	*GPT4+XGB	0.55	0.33	0.32	0.3
5	0.5	DANets	**0.58**	**0.49**	**0.65**	**0.48**
	–	*GPT4+1NN	0.47	0.32	0.29	0.3
36	No-Fusion	*1NN	0.27	0.18	0.19	0.18

Table 1: Best performances (on F1) of baselines and FuSE-MET with the classifiers. '*' indicates baseline.

We trained multiple classifiers, including Random Forest (RF), 1-Nearest Neighbor (1NN), SVM, Fully-connected Neural Networks (NN), XGB, DANets, and CNN, using the merged labels generated by FuSE-MET. FuSE-MET was compared to two baselines: (i) GPT-4-based label fusion without MET values and (ii) a nonfusion model with 36 distinct clustering. Fig. 2 shows that FuSE-MET consistently outperforms the baselines by effectively merging labels using both semantic meaning and MET values. Optimal performance is achieved with λ in the range of 0.2-0.5, balancing domain knowledge and semantics (Fig. 3). Lower k values usually provide higher accuracy but K=3 has a higher F1 (0.77) score than that of K=2 (0.65), which indicates K=3 is a more accurate hyperparameter for label fusion than K=2. (Table 1). Additionally, simpler models (i.e., 1NN, SVM) perform better as FuSE-MET's reduced label complexity allows them to efficiently capture patterns without overfitting.

4 Conclusion

FuSE-MET shows superior performance in activity recognition by merging labels by leveraging semantic meaning and MET values; Our approach reduces label complexity, enabling more accurate classification of sensor data. Classifiers trained on the optimized labels generated by FuSE-MET consistently outperform the baselines, including a ChatGPT-4-based label fusion without domain knowledge and a no-fusion model with distinct clustering. By balancing semantic information and physical intensity, FuSE-MET proves to be a robust and scalable solution for real-world activity recognition tasks, especially in uncontrolled environments.

5 Acknowledgments

This work was supported in part by the National Institutes of Health under grant 1R21NR015410-01 and the National Science Foundation under grant CNS-2227002. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.

References

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *North American Chapter of the Association for Computational Linguistics*.

Ermes, M.; Pärkkä, J.; Mäntyjärvi, J.; and Korhonen, I. 2008. Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions. *IEEE Transactions on Information Technology in Biomedicine*, 12: 20–26.

Fullerton, E.; Heller, B. W.; and Muñoz-Organero, M. 2017. Recognizing Human Activity in Free-Living Using Multiple Body-Worn Accelerometers. *IEEE Sensors Journal*, 17: 5290–5297.

Mirzadeh, S. I.; Ardo, J.; Fallahzadeh, R.; Minor, B.; Evangelista, L.; Cook, D.; and Ghasemzadeh, H. 2019. LabelMerger: Learning Activities in Uncontrolled Environments. In *2019 First International Conference on Transdisciplinary AI (TransAI)*, 64–67.