



HAL
open science

Representation of Irregularly Sampled Time Series with Generative Language Models for Classification and Transfer Learning

Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit Leduc, Ioannis Kanellos

► To cite this version:

Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit Leduc, Ioannis Kanellos. Representation of Irregularly Sampled Time Series with Generative Language Models for Classification and Transfer Learning: a Case Study in Activities of Daily Living Recognition. ML4ITS 2023 - Machine Learning for Irregular Time Series Workshop in ECML PKDD, Sep 2023, Turin (IT), Italy. hal-04437409

HAL Id: hal-04437409

<https://hal.science/hal-04437409>

Submitted on 4 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Representation of Irregularly Sampled Time Series with Generative Language Models for Classification and Transfer Learning: a Case Study in Activities of Daily Living Recognition

Damien Bouchabou^{1,2}[0000-0003-3623-3626], Sao Mai Nguyen^{1,2}[0000-0003-0929-0019], Christophe Lohr²[0000-0002-0655-2880], Benoit LeDuc³, and Ioannis Kanellos²[0000-0001-5323-1601]

¹ U2IS, ENSTA, IP Paris, France

{damien.bouchabou,nguyensmai}@gmail.com

² RAMBO, IMT Atlantique Engineer School, 29238 Brest, France

{christophe.lohr,ioannis.kanellos}@imt-atlantique.fr

³ Delta Dore Company, 35270 Bonnemain, France bleduc@deltadore.com

Abstract. Human Activity Recognition plays a crucial role in various fields, including healthcare and smart homes. With the increasing prevalence of smart homes equipped with ambient sensors, there is a growing interest in leveraging artificial intelligence techniques to understand and recognize human activities within these environments. However, the irregular and noisy nature of data collected by ambient sensors presents unique challenges. To address these challenges, we propose the use of a pre-trained embedding trained on sensor activation sequences, specifically an algorithm based on an architecture similar to GPT, to improve the classification performance of activities of daily living in smart homes. Additionally, we investigate the concept of transfer learning between smart homes, utilizing the knowledge gained from one environment to enhance activity recognition in another. The results demonstrate that the GPT transformer decoder approach outperforms other algorithms in terms of accuracy and balanced accuracy across multiple datasets. The findings also highlight the potential of transfer learning, with GPT transformer decoder pre-trained embeddings from a clean and large dataset showing promising results in various scenarios.

Keywords: Human Activity Recognition · Activity of Daily Living · Smart Homes · Transfer Learning · Generative Models · Self-Supervised Learning.

1 Introduction

Human activity recognition (HAR) and in particular Activities of Daily Living (ADL) recognition focuses on algorithms capable of finding patterns in smart

homes equipped with ambient sensors (eg. motion, door open/close, temperature) so as to provide services for healthcare or enhance daily life.

In smart homes, ambient sensors collect data that form a multi variate time series. The data are event-triggered by sensors that have different response timings and have imperfect accuracy, thus the time series is noisy and irregularly sampled. This makes it challenging to model using traditional time series methods. Moreover, ambient sensors give little information on the context and state of the environment. While this enable them to be considered as respecting privacy, this also results in sparse information, that do not hold the Markov property [2]. Deep learning models [13, 1], on the other hand, have shown promising results in handling irregular time series data due to their ability to learn complex temporal patterns from large datasets and representing contextual sequences. However, only few datasets of HAR from smart homes exist because of difficulty in labelling, while data from smart homes vary considerably with intra- and inter-variability of activity execution because of changes of habits, household composition, house architecture and IoT network facility.

The research in HAR involves developing novel deep learning models that can effectively model irregular time series data and be transferred to new households with (1) Sequence models for modeling irregular time series data such as Recurrent neural networks and their variants such as bi-LSTM [9] or language models such as ELMo [10] trained on sensor activation sequences [3], these algorithms have leveraged pre-trained embeddings by transforming time series data into symbols, to enhance the classification performance of these sequences into ADLs; (2) Attention mechanisms that can dynamically weigh the importance of different time steps in long sequences.

Combining both ideas of using sequence models based on generative language models and attention models, we propose a small GPT-architecture model [11] trained on sensor activation sequences to be used as a pre-trained embedding. This pre-trained embedding aims to represent the multivariate irregular time series by enriching each sensor activation log by its context. We show that this pre-trained embedding allows better classification of ADL, as well as transfer learning between smart homes: this representation leverages the knowledge gained from one smart home environment to improve activity recognition in another.

In this article, for HAR using ambient sensors, we explore two key approaches: leveraging the generalization power and transfer-learning abilities of pre-trained embeddings, and harnessing the capabilities of transformer-based architectures to capture contextual and temporal information. To achieve this, we will utilize a pre-trained GPT transformer decoder embedding trained on sensor activations for human activity recognition within smart homes. Additionally, we investigate the potential of transfer learning between different smart home environments. We report the results for both studies on three datasets, varying in noise levels.

Our contribution is to illustrate with the case of HAR using ambient sensors, how deep learning using pre-trained embeddings with transformer capture contextual and temporal information of evenly sampled multivariate time series for classification and transfer learning, and test its performance for noisy data.

2 Related Works

Event-triggered Time Series Recognition As an event-triggered sequence of multi-modal data, recognizing ADLs from a network of IoT sensors poses the problem of noisy and irregular time series classification. Various methods have been studied for recognizing ADLs in smart homes. Early approaches, such as Support Vector Machines (SVM), Naive Bayes networks, and Hidden Markov Models (HMM) [12], lack generalization and adaptability due to their reliance on hand-crafted features and specific contexts. Recently, deep learning techniques, particularly Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), have shown promise in sequence modeling.

CNNs excel at feature extraction and pattern recognition, capturing local dependencies and handling differences in step frequency or event occurrence. For example, Gochoo et al. [7] transformed activity sequences into binary images to leverage 2D CNN-based structures. Singh et al. [13] applied a 1D CNN structure to raw data sequences, demonstrating its high feature extraction capability. Bouchabou et al. [1] also utilized a 1D CNN structure inspired by Fully Convolutional Networks, achieving competitive results compared to LSTM structures. However, LSTM-based models still outperform the CNN 1D architecture.

LSTM models are specifically designed to handle time sequences and effectively capture both long and short-term dependencies. In the context of HAR in smart homes, Liciotti et al. [9] extensively investigated various LSTM structures and demonstrated their superiority over traditional HAR approaches in terms of classification scores without the need for handcrafted features. LSTM's ability to generate features that encode temporal patterns has been highlighted in [14] when compared to conventional machine learning techniques. As a result, LSTM-based structures have emerged as the leading models for tackling the challenges of HAR in the smart home domain.

Transfer Learning Building a HAR model for each smart home is impractical due to variations in floor plans and sensor configurations, making it challenging to acquire labeled data for each home. This limitation stems from cost and privacy concerns. To address this issue, transfer learning has emerged as a solution, enabling the adaptation to new home environments [4]. Hu et al. [8] proposed a transfer learning framework by automatically learning a correspondence between different sets of sensors and demonstrated its effectiveness over two datasets. Yu et al. [15] introduced a novel approach by transforming activity sequences from a dataset into vectors using a Word2Vec model. They then fine-tuned the Word2Vec model on a new dataset, showcasing competitive performances. Recently, Bouchabou et al. [3] leveraged advanced LSTM structures commonly used in NLP to enhance human activity recognition. Inspired by the ELMo model [10], their proposed LSTM structure utilizes a pre-trained embedding and has demonstrated superior performance compared to other methods. Furthermore, it showcases the ability to leverage transfer learning between different datasets. Our approach uses a transformer-based embedding instead.

Our approach builds upon the methods described in [3]. We employ a self-supervised sensors embedding on each dataset, leveraging the power of pre-trained embeddings such as ELMo [10] and GPT [11]. The learned sensors embeddings are then utilized for classification.

3 Approach

In this section, we describe the datasets used and the proposed algorithm.

3.1 Datasets

To evaluate the performance and effectiveness of our approach, we utilize several benchmark datasets commonly used in the field of HAR via ambient sensors in smart homes. Specifically, we employ the Aruba, Milan, and Cairo datasets from the CASAS [6] benchmark dataset collection.

Each of these datasets covers a diverse range of activities and captures the complexity of real-world human behavior within smart home environments. The selected datasets offer different characteristics in terms of the house structure and the number of inhabitants, as summarized in Table 1.

Table 1. Dataset Details

Dataset	Aruba	Milan	Cairo
Residents	1	1 + pet	2 + pet
Number of Sensors	39	33	27
Number of Activities	12	16	13
Number of Days	219	82	56

These datasets include several months of labeled activities and are unbalanced. The inclusion of unbalanced data provides a realistic representation of real-world scenarios, where some activities may occur less frequently.

The selection of these three datasets allows us to demonstrate our approach in different scenarios: (1) a single inhabitant as a baseline (Aruba), (2) a more complex situation with the presence of a pet introducing noise (Milan), and (3) a multi-resident scenario with potential interferences from another resident (Cairo). In the case of Milan and Cairo, sensor activations triggered by pets or other residents may introduce perturbations or noise in the data. Sensor activations triggered by pets are typically less disruptive than those triggered by human residents, as sensors are primarily designed to detect human activities.

By utilizing these benchmark datasets, we can thoroughly evaluate the performance of our approach and assess its resilience to noise, interferences, and unbalanced data distributions.

3.2 Architecture

Our proposed architecture is inspired by the work of Liciotti et al. [9] and Bouchabou et al. [3]. Similar to their approach, we utilize a deep learning model that

includes an embedding or a pre-trained embedding layer followed by a classification head. The classification head consists of a bi-LSTM layer and a Softmax layer, as illustrated in Fig.1. In our implementation, we employ a GPT transformer decoder model as the pre-trained embedding.

We have developed a simplified version of the renowned GPT model. Our model incorporates 3 decoder transformer layers with an embedding dimension of 384 and 8 attention heads.

In contrast to the original GPT model’s usage of the Gelu activation function, we have opted for the Relu activation function. Through experiments, we have observed that the Relu activation yields superior results in our case. However, similar to the original GPT model, we apply pre-normalization, i.e. normalization to the inputs of the attention layer(s) and the intermediate dense layer.

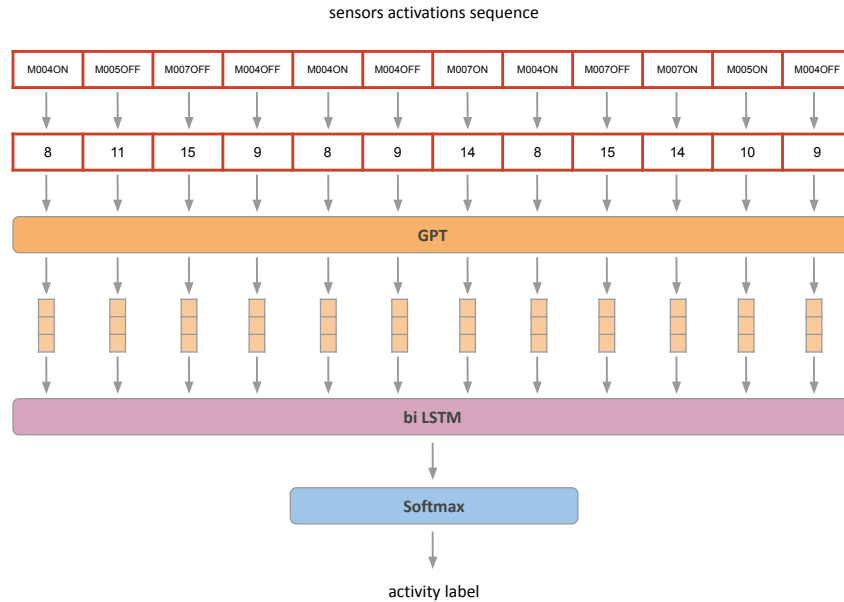


Fig. 1. Proposed architecture

3.3 Training and Evaluation Methodology

We applied the same methodology across all datasets to thoroughly evaluate our approach. Three models were trained: bi-LSTM [9], ELMo [3], and our GPT model.

Data Pre-processing: Each dataset is divided into weeks to maintain temporal coherence. Weeks are further split into training and testing sets. Original

activity labels are retained without renaming or grouping them into meta-labels, in contrast to previous studies such as [3, 9, 5]. This introduces challenges due to similar sensor activation patterns among activities, especially when they occur in the same area (e.g., breakfast, lunch, dining, washing dishes, etc.).

Data Encoding: We employ the method proposed by Bouchabou et al. [3] to encode the datasets. Sensor events are converted into words by concatenating the sensor ID s_i and value v_i (e.g., $s_i = \text{M001}$ and $v_i = \text{ON}$ becomes M001ON). Real-valued sensors are represented in a similar manner. This approach allows us to treat numerical values as categorical values. Each unique word contributes to the smart home vocabulary, which describes activities.

Similar to NLP, each event in the stream is transformed into an index that can be used by a neural network. Indexing begins at 1, reserving 0 for sequence padding. Indexes are assigned based on word frequency, with higher-frequency words receiving lower index values. The inputs models consist of sequences of index representing sensor activations, maintaining the order of timestamps (Fig.1).

Training Pre-trained Embeddings: We trained two sensor embeddings using the training set in two distinct ways. For ELMo embedding, we divided each week into segments representing individual activities to capture context within each sequence as the original ELMo model [10].

For GPT embedding, we divided each week into overlapping windows of a fixed size (e.g., 1024 sensor events) with a specified overlap (e.g., 512 sensor events). The GPT model was trained with sliding windows instead of individual activity sequences to learn a global context and structural representation of sensor activations, similar to the original GPT model. It’s important to note that each pre-trained embedding was prepared using a single dataset.

Classifier Training: To prepare the different classifier algorithms, we used the training set of each dataset and employed a 3-fold cross-validation method. For models with a pre-trained embedding, the embeddings were frozen, and only the classification head was trained. Models were trained using pre-segmented activity sequences.

Evaluation: Finally, to assess the generalization ability and effectiveness of the trained models, we utilized the testing set of each dataset. Only the best model provided by the 3-fold cross-validation method was evaluated for each model.

By following this comprehensive training and evaluation methodology, we aim a robust assessment of our approach’s performance on the tested datasets.

4 Results

4.1 Activity of Daily Living Recognition

In this subsection, we report the classification results obtained from the three algorithms: Liciotti et al. [9], ELMo by Bouchabou et al. [3], and our proposed GPT-transformer-decoder approach. We evaluate the performance of these algorithms using both 3-fold cross-validation and test sets of three CASAS datasets [6]: Aruba, Milan, and Cairo. The cross-validation results are summarized in Table 2, while the results on the test datasets are presented in Table 3.

Cross-Validations Results: Table 2 reports the average accuracy and balanced accuracy obtained through 3-fold cross-validation for each algorithm and dataset.

Across all datasets, all algorithms demonstrate high accuracy rates, indicating their effectiveness in activity recognition. Liciotti et al.’s approach achieves competitive accuracy, with an accuracy of 96.10% attained in the Aruba dataset. Our GPT-transformer-decoder approach is slightly above with an accuracy of 96.52%, while ELMo method achieves the highest accuracy of 96.79%. For the Milan dataset, our GPT-transformer-decoder approach outperforms the other algorithms, achieving an accuracy of 88.47%. ELMo method follows closely with an accuracy of 87.85%, while Liciotti et al.’s approach achieves an accuracy of 86.97%. For the Cairo dataset, our GPT-transformer-decoder approach demonstrates its superiority with an accuracy of 83.12%. The ELMo method achieves an accuracy of 79.40%, and Liciotti et al.’s approach achieves an accuracy of 72.61%.

Considering the balanced accuracy metric, our GPT-transformer-decoder approach consistently outperforms the other algorithms across all datasets. This indicates its ability to handle imbalanced activity classes effectively and make accurate predictions for both major and minor classes.

Table 2. Comparison of Liciotti et al. ELMo and GPT (averaged over 3-fold cross-validation)

	Aruba		Milan		Cairo	
	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
Liciotti et al.	96.10%	80.71%	86.97%	67.70%	72.61%	52.52%
ELMo (Bouchabou et al.)	96.79%	80.70%	87.85%	69.38%	79.40%	65.84%
GPT (ours)	96.52%	81.36%	88.47%	73.47%	83.12%	72.18%

Test Sets Results: Table 3 presents the evaluation on the test sets for each algorithm on the Aruba, Milan, and Cairo datasets test sets to assess their generalization performance.

In the Aruba dataset, our GPT-transformer-decoder approach achieves the highest accuracy of 97.10%, surpassing both Liciotti et al.’s approach with 96.31% accuracy and ELMo method with 96.91% accuracy. The balanced accuracy metric also favors our GPT-transformer-decoder approach with an accuracy of 85.53%. For the Milan dataset, our GPT-transformer-decoder approach maintains its superior performance with an accuracy of 87.71%. The ELMo method follows closely with 86.79%, while Liciotti et al.’s approach achieves 85.98%. Once again, our GPT-transformer-decoder approach exhibits a higher balanced accuracy of 67.38%. In the Cairo dataset, our GPT-transformer-decoder approach continues to outperform the other algorithms, achieving an accuracy of 81.89%. The ELMo method achieves an accuracy of 76.45%, and Liciotti et al.’s approach achieves an accuracy of 72.74%. The balanced accuracy also favors our GPT-transformer-decoder approach with an accuracy of 71.49%.

Our GPT-transformer-decoder approach demonstrates superior performance in terms of balanced accuracy compared to other algorithms, indicating its effectiveness in handling imbalanced datasets. This highlights the robustness of our GPT-transformer-decoder model in accurately recognizing activities.

Furthermore, the results obtained from the test datasets provide further validation of the effectiveness of our GPT-transformer-decoder approach for activity recognition in smart homes. It consistently achieves high accuracy rates, surpassing the other algorithms in both accuracy and balanced accuracy metrics.

Table 3. Comparison of Liciotti et al. ELMo and GPT (test set)

	Aruba		Milan		Cairo	
	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
Liciotti et al.	96.31%	84.49%	85.98%	66.10%	71.74%	56.94%
ELMo (Bouchabou et al.)	96.91%	84.14%	86.79%	66.69%	76.45%	63.33%
GPT (ours)	97.10%	85.53%	87.71%	67.38%	81.89%	71.49%

Discussion: Our GPT-transformer-decoder approach consistently outperforms other algorithms in terms of accuracy and balanced accuracy, as demonstrated in both cross-validation and test set evaluations. This remarkable performance can be attributed to the transformer architecture’s ability to capture intricate patterns and dependencies within sensor activations during the self-supervised task.

Furthermore, our GPT-transformer-decoder approach consistently exhibits higher balanced accuracy across all datasets, indicating its robustness in handling imbalanced data. This is particularly relevant in smart home environments where certain activities occur less frequently than others. Our approach effectively addresses this challenge by providing reliable predictions for both major and minor activity classes.

The presence of pets in the Milan dataset introduces additional complexity, as they introduce noise in sensor activations and activity patterns. However, our GPT-transformer-decoder approach maintains its superior performance even in the presence of pets, highlighting its adaptability to diverse smart home scenarios.

In the multi-resident scenario of the Cairo dataset, potential interferences caused by another resident pose additional challenges. While the accuracy and balanced accuracy of the other algorithms drop by more than 10% compared to the Milan dataset, our GPT-transformer-decoder approach’s accuracy only decreases by about 6% while the balanced accuracy even increases by about 4%. This suggests that the GPT-transformer-decoder model can effectively handle interferences caused by multiple residents, enabling accurate recognition of activities in such settings.

The superior performance of our approach on noisy and interfered datasets can be attributed to the self-supervised task learning. During this task, the model learns to predict the next sensor activation based on the previous context, gaining an understanding of the logic behind sensor activations and their order. This enables our model to effectively mitigate noise and interference.

In conclusion, our results highlight the effectiveness of our GPT-transformer-decoder approach for recognizing activities of daily living in smart home environments. With its competitive performance, ability to handle class imbalances, mitigate noise and interference, and adapt to diverse scenarios, our approach shows great promise as a reliable solution for classification of irregular time series.

4.2 Transfer Learning Between Smart Homes:

In this sub-section, we analyze and compare the performance of ELMo and our GPT-transformer-decoder algorithms for transfer learning in human activity recognition via ambient sensors in smart homes, as demonstrated by Bouchabou et al. [3]. We investigate the effectiveness of transferring knowledge between different datasets. Firstly, we evaluate the performance of ELMo and our GPT-transformer-decoder models with and without pre-trained embeddings from the clean and large Aruba dataset. Then, we assess the transferability of the medium-sized and noisy Milan dataset. The goal is to understand the potential of transfer learning in activity recognition across diverse smart home environments.

Transfer Learning from Aruba to Milan and Cairo Datasets: We assess the transfer learning capabilities by comparing the classification results on the Milan and Cairo datasets with and without utilizing embeddings trained on the Aruba dataset. The evaluation metrics, including accuracy and balanced accuracy, are presented in Table 4. The results are averaged over 3-fold cross-validation for different scenarios.

In the Milan dataset, both ELMo and our GPT-transformer-decoder models demonstrate promising results with transfer learning. ELMo achieves an accuracy of 87.16% and a balanced accuracy of 69.02% with the Aruba-trained

Table 4. Comparison of ELMo and GPT pre-trained embeddings, for transfer learning from Aruba (averaged over 3-fold cross- validation)

	Milan		Milan with Embedding trained on Aruba		Cairo		Cairo with Embedding trained on Aruba	
	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
Liciotti et al.	86.97%	67.70%	NA	NA	72.61%	52.52%	NA	NA
ELMo (Bouchabou et al.)	87.85%	69.38%	87.16%	69.02%	79.40%	65.84%	76.68%	61.73%
GPT (Ours)	88.47%	73.47%	89.59%	71.40%	83.12%	72.18%	82.69%	67.40%

embedding, slightly lower than the ELMo model without transfer learning but still better than the baseline Liciotti et al. model.

Our GPT-transformer-decoder model exhibits improved performance with transfer learning. The accuracy increases from 88.47% to 89.59%, indicating the benefits of leveraging pre-trained embeddings from the Aruba dataset for activity recognition in the Milan dataset. Although the balanced accuracy decreases from 73.47% to 71.40%, our GPT-transformer-decoder model with transfer learning remains competitive and outperforms both the Liciotti et al. baseline model and ELMo without transfer learning.

Similar observations can be made for the Cairo dataset. Both ELMo and our GPT-transformer-decoder models show slightly lower performance with transfer learning, but still outperform the baseline Liciotti et al. model. ELMo achieves an accuracy of 76.68% and a balanced accuracy of 61.73% with Aruba-trained embeddings, while our GPT-transformer-decoder model achieves an accuracy of 82.69% and a balanced accuracy of 67.40%. Once again, our GPT-transformer-decoder model outperforms the ELMo model in this task.

Table 5. Comparison of ELMo and GPT pre-trained embeddings, for transfer learning from Aruba (test sets)

	Milan		Milan with Embedding trained on Aruba		Cairo		Cairo with Embedding trained on Aruba	
	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
Liciotti et al.	85.98%	66.10%	NA	NA	71.74%	56.94%	NA	NA
ELMo (Bouchabou et al.)	86.79%	66.69%	85.63%	66.03%	76.45%	63.33%	77.26%	61.73%
GPT (Ours)	87.71%	67.38%	86.84%	63.16%	81.89%	71.49%	79.88%	62.74%

Table 5 compares the performance of ELMo and our GPT-transformer-decoder models on the test sets of the Milan and Cairo datasets, evaluating their generalization ability.

For the Milan dataset, our GPT-transformer-decoder approach with Aruba embeddings achieves a superior accuracy of 86.84% compared to ELMo and Liciotti et al. approaches. However, the GPT-transformer-decoder approach has the lowest balanced accuracy. The ELMo method performs similarly to its non-transfer learning version, with an accuracy of 85.63% and a balanced accuracy of

66.03%. Both models demonstrate high performance, but transfer learning does not enhance the classification results on this test set.

In the Cairo dataset, our GPT-transformer-decoder approach with transfer learning outperforms the other algorithms, achieving higher accuracy and better balanced accuracy than ELMo with transfer learning. However, the transferred embeddings still exhibit better performance than the Liciotti et al. baseline but lower performance than their own version without transfer learning embeddings.

This experiment demonstrates that algorithms with embeddings trained from a larger dataset can provide high classification performance, although lower than without transfer learning.

Transfer from Milan to Aruba and Cairo Datasets: To further evaluate the transfer learning capabilities, we examine the performance of ELMo and our GPT models trained on the Milan dataset and tested on the Aruba and Cairo datasets. The results are presented in Table 5.

Table 6. Comparison of ELMo and GPT pre-trained embeddings, for transfer learning from Milan (averaged over 3-fold cross-validation)

	Aruba		Aruba with Embedding trained on Milan		Cairo		Cairo with Embedding trained on Milan	
	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
Liciotti et al.	96.10%	80.71%	NA	NA	72.61%	52.52%	NA	NA
ELMo (Bouchabou et al.)	96.79%	80.70%	94.87%	74.87%	79.40%	65.84%	79.11%	63.57%
GPT (Ours)	96.52%	81.36%	94.88%	70.98%	83.12%	72.18%	81.69%	68.16%

For the Aruba dataset, both ELMo and our GPT-transformer-decoder models with transfer learning demonstrate competitive performance. However, the models with transferred embeddings exhibit lower performance compared to their counterparts without transfer learning. Additionally, our GPT-transformer-decoder model with transfer learning shows poorer balanced accuracy compared to the ELMo model. It is important to mention that the transferred models also perform worse than the Liciotti et al. baseline models, indicating lower performance on the Aruba test set.

Regarding the Cairo dataset, the GPT-transformer-decoder approach with transfer learning outperforms both the Liciotti et al. and ELMo models, regardless of transfer learning. Similar to the Aruba dataset, the GPT-transformer-decoder model with transfer learning achieves lower performance compared to its counterpart.

The evaluation of the test sets, as shown in Table 7, yields similar results to the cross-validation results. Models with transfer learning demonstrate competitive performance but fall short compared to their counterparts. Notably, our GPT-transformer-decoder model consistently outperforms the others on the Cairo dataset.

Table 7. Comparison of ELMo and GPT pre-trained embeddings, for transfer learning from Milan (test sets)

	Aruba		Aruba with Embedding trained on Milan		Cairo		Cairo with Embedding trained on Milan	
	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
Liciotti et al.	96.31%	84.49%	NA	NA	71.74%	56.94%	NA	NA
ELMo (Bouchabou et al.)	96.91%	84.14%	94.62%	77.19%	76.45%	63.33%	74.85%	62.45%
GPT (Ours)	97.10%	85.53%	94.70%	75.62%	81.89%	71.49%	78.67%	66.22%

Discussion: The transfer learning experiments confirm the findings of Bouchabou et al. [3], showing that pre-trained embeddings can enhance classification performance. Larger datasets, such as transferring from Aruba to Milan and Cairo, yield better performance compared to medium-sized and noisy datasets, as observed in the transfer from Milan to Aruba and Cairo. Even with medium-sized and noisy datasets, reasonably high levels of accuracy and balanced accuracy are achieved.

Notably, embeddings trained on the Aruba dataset produce comparable results to pre-trained embeddings without transfer. This suggests that pre-training the embedding on the target dataset may not be necessary, especially when the pre-trained embedding originates from a large and clean dataset like Aruba. The transfer of knowledge from the Aruba dataset already provides promising results, indicating that the pre-trained embedding captures relevant features that generalize well to different datasets.

The success of transfer learning can be attributed to the shared sensor denominations across the datasets. Despite using different sensor IDs for the same sensors, the datasets share the same underlying vocabulary in terms of sensor types. Although consistent sensor IDs are not used across datasets, the transfer learning approach remains effective because the pre-trained embedding generates similar features for sensor sequences belonging to the same activity class. The classifier head can then utilize these features to associate them with the correct activity labels.

An analogy can be made with NLP, where pre-trained embeddings are trained on a specific language and the classifier operates on a dialect version of that language. The role of the classifier is to establish the parallel between the language and the dialect, effectively translating features for the final task.

Consistently, our observations show that the GPT embedding outperforms the ELMo embedding in the transfer learning task. This trend is particularly evident in the Cairo dataset, which involves multiple residents, despite the pre-trained dataset (Aruba) containing data from a single resident. This suggests that the GPT embedding exhibits stronger adaptability and generalization capabilities, enabling it to effectively handle the complexities arising from multi-resident scenarios.

5 Discussion and Conclusion

We have explored the use of generative pre-training models, specifically a GPT-based architecture, for evenly spaced and noisy multivariate time series through

the use case of human activity recognition via ambient sensors in smart homes. We have investigated the performance of our proposed approach in terms of ADL recognition and transfer learning between different smart home environments.

Our results demonstrate that the GPT-transformer-decoder approach outperforms other algorithms, including the bi-LSTM-based approach by Liciotti et al. [9] and the ELMo-based approach by Bouchabou et al. [3], in terms of accuracy and balanced accuracy across different datasets. This highlights the effectiveness of transformer-based architectures in capturing intricate patterns and dependencies within sensor activations. We have also explored the potential of transfer learning using pre-trained embeddings. Our experiments show that pre-trained embeddings trained on one dataset can be transferred to other datasets, resulting in improved classification performance. Transferring knowledge from a larger and cleaner dataset to smaller and noisier datasets yields better results. However, even with medium-sized and noisy datasets, transfer learning with pre-trained embeddings achieves reasonably high levels of accuracy and balanced accuracy. The success of transfer learning can be attributed to the shared sensor denominations across datasets, allowing pre-trained embeddings to capture relevant features that generalize well. This suggests that pre-training on a larger and cleaner dataset is sufficient for effective transfer learning, eliminating the need for pre-training on the target dataset.

In conclusion, our study indicates the potential of our GPT-transformer-decoder model for classification of irregularly sampled and noisy time series. In our case study, our approach exhibits superior performance compared to other algorithms, enabling accurate recognition of time series in diverse noise levels. We have confirmed the transfer learning ability of pre-trained embeddings, as demonstrated by Bouchabou et al. [3], allowing knowledge transfer between different smart home environments. This transfer learning capability enhances classification performance in scenarios with limited labeled data, data heterogeneity, and noise. Notably, our GPT-transformer-decoder approach outperforms other algorithms in this task.

Future work can focus on exploring larger and more diverse datasets, investigating different architectures and configurations of GPT-like models, leveraging the ability of transformers to consider a broader context, and integrating timestamps information. These directions can enhance the performance and understanding of GPT-transformer-decoder models for activity recognition in smart homes. Overall, the use of generative pre-training models and transfer learning opens up possibilities for enhancing smart home efficiency, improving healthcare services, and enabling personalized assistance in smart home environments.

Acknowledgment

This work is partially supported by project VITAAL financed by Brest Metropole, the region of Brittany and the European Regional Development Fund (ERDF); and partially supported by the “plan France relance” of December, 21th, 2020.

References

1. Bouchabou, D., Nguyen, S.M., Lohr, C., LeDuc, B., Kanellos, I.: Fully convolutional network bootstrapped by word encoding and embedding for activity recognition in smart homes. In: Li, X., Wu, M., Chen, Z., Zhang, L. (eds.) *Deep Learning for Human Activity Recognition*. pp. 111–125. Springer Singapore (February 2021)
2. Bouchabou, D., Nguyen, S.M., Lohr, C., LeDuc, B., Kanellos, I.: A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning. *Sensors* **21**(18), 6037 (sep 2021)
3. Bouchabou, D., Nguyen, S.M., Lohr, C., LeDuc, B., Kanellos, I.: Using language model to bootstrap human activity recognition ambient sensors based in smart homes. *Electronics* **10**(20), 2498 (2021)
4. Cook, D., Feuz, K.D., Krishnan, N.C.: Transfer learning for activity recognition: A survey. *Knowledge and information systems* **36**, 537–556 (2013)
5. Cook, D.J.: Learning setting-generalized activity models for smart spaces. *IEEE intelligent systems* **2010**(99), 1 (2010)
6. Cook, D.J., Crandall, A.S., Thomas, B.L., Krishnan, N.C.: Casas: A smart home in a box. *Computer* **46**(7), 62–69 (2012)
7. Gochoo, M., Tan, T.H., Liu, S.H., Jean, F.R., Alnajjar, F.S., Huang, S.C.: Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and dcnn. *IEEE journal of biomedical and health informatics* **23**(2) (2018)
8. Hu, D., Yang, Q.: Transfer learning for activity recognition via sensor mapping. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain*. p. 1962 (2011)
9. Liciotti, D., Bernardini, M., Romeo, L., Frontoni, E.: A sequential deep learning application for recognising human activities in smart homes. *Neurocomputing* (04 2019). <https://doi.org/10.1016/j.neucom.2018.10.104>
10. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*. vol. 1, pp. 2227–2237 (Jun 2018). <https://doi.org/10.18653/v1/N18-1202>, <https://aclanthology.org/N18-1202>
11. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I.: Improving language understanding by generative pre-training. In: *Technical Report* (2018)
12. Sedky, M., Howard, C., Alshammari, T., Alshammari, N.: Evaluating machine learning techniques for activity classification in smart home environments. *International Journal of Information Systems and Computer Sciences* **12**(2), 48–54 (2018)
13. Singh, D., Merdivan, E., Hanke, S., Kropf, J., Geist, M., Holzinger, A.: Convolutional and recurrent neural networks for activity recognition in smart environment. In: *Towards integrative machine learning and knowledge extraction*, pp. 194–205. Springer (2017)
14. Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M., Holzinger, A.: Human activity recognition using recurrent neural networks. In: *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. pp. 267–274. Springer (2017)
15. Yu, Y., Tang, K., Liu, Y.: A fine-tuning based approach for daily activity recognition between smart homes. *Applied Sciences* **13**(9), 5706 (2023)