Poster Abstract: Fine-grained Contextualized Activity Logs Generation based on Multi-Modal Sensor Data and LLM

Ye Tian, Onat Gungor, Xiaofan Yu, Tajana Rosing
Computer Science and Engineering
University of California San Diego
La Jolla, California, USA
{yet002,ogungor,x1yu,tajana}@ucsd.edu

ABSTRACT

Detailed activity logs are crucial for health monitoring and personalized interventions. Traditional methods rely on manual editing or raise privacy concerns due to the use of camera recordings. This paper proposes ContextLLM, an innovative system that utilizes a large language model (LLM) to understand sensor data from smartphones and smartwatches and automatically generate contextualized activity logs. Compared to the state-of-the-art, it incorporates key contextual information and physiological indicators, enabling more fine-grained semantic descriptions. Preliminary results show that the automatically generated activity logs achieve 80.26% similarity to human annotations, demonstrating the feasibility.

KEYWORDS

Cyber-physical systems, activity logs, large language model

ACM Reference Format:
Ye Tian, Onat Gungor, Xiaofan Yu, Tajana Rosing. 2025. Poster Abstract: Fine-grained Contextualized Activity Logs Generation based on Multi-Modal Sensor Data and LLM. In *The 23rd International Conference on Information Processing in Sensor Networks (SenSys '25), May 6–9, 2025, Irvine, USA.* ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

In 2024, global wearable device shipments surpassed 1.2 billion units. These rapid growth wearable devices is providing rich sensing data for daily activity recognition and health monitoring. However, most existing studies on human activity recognition primarily focus on classifying basic movements [4], with limited analysis of contextual factors such as location, environment, and physiological changes. Fine-grained contextual records of activities can contribute to daily health monitoring and timely personalized interventions, especially for elderly individuals with cognitive decline. Additionally, detailed contextualized activity logs also provide users with richer information, enabling them to better review their daily lives and health status [7]. Currently, some commercial applications facilitate activity log generation [2], but they rely on manual data uploads and textual annotations, imposing a significant user burden and hindering systematic, consistent logging. Some studies investigate wearable cameras and smart glasses for daily life logging [1], but these devices are costly and pose privacy concerns. Recently, Autolife [7] has leveraged 4 types of low-cost smartphone sensor data to autonomously generate life logs. By integrating map images and WiFi Service Set Identifiers (SSIDs), it employs LLMs to generate detailed activity location descriptions and subsequently summarize the user's life logs. However, crucial environmental context and physiological changes during activities are not accounted for. Additionally, AutoLife trains four Large Language Models (LLMs) to analyze and aggregate spatial context information, leading to a highly complex processing workflow. In contrast to prior work, we propose an innovative system, ContextLLM, for generating activity logs with fine-grained contextual information while optimizing the efficiency of the inference workflow.

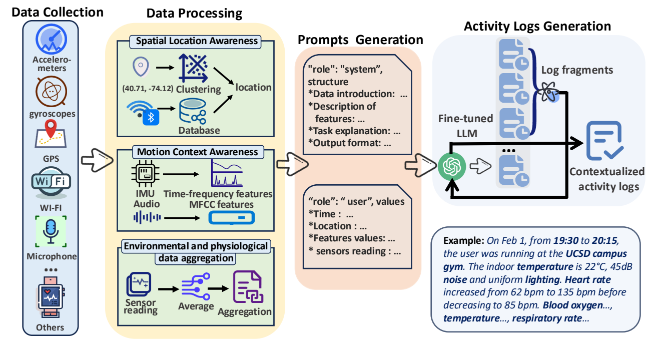

Figure 1: ContextLLM overview

Spatial Context Awareness Sensors	GPS, Barometer, Wi-Fi and Bluetooth.
Motion State Monitoring Sensors	Accelerometers, Gyroscopes and Magnetometers
Environmental Parameter Detection Sensors	Temperature, Microphone and Light sensors.
Physiological Indicator Monitoring Sensors	PPG, SpO2, GSR/EDA, ECG, Thermometers.

Table 1: Categorizing sensors on smartphones and watches

Developing ContextLLM presented two key challenges: 1) integrating multi-modal, heterogeneous sensor data and accurately interpreting its meaning, and 2) efficiently generating activity logs with precise semantic descriptions. To address them, we categorize the sensors into four major groups based on the type of information they represent, as shown in Table 1. Then, we analyze the structure of different sensor data and design specific feature extraction algorithms to process the data. Subsequently, we carefully design a prompt generator to dynamically transform the extracted multi-modal features into structured text and fine-tune the LLM to understand and interpret sensor data. These important method designs significantly enhance the LLM's reasoning ability for long time-series data. Finally, we align and merge the output descriptions from continuous time windows based on semantic information to generate fine-grained contextualized activity logs.

SenSys '25, May 6–9, 2025, Irvine, USA

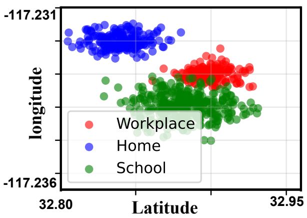

Figure 2: Location Clustering

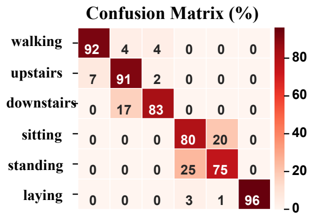

Figure 3: User's Motions

2 SYSTEM DESIGN

Fig. 1 presents an overview of ContextLLM . It collects sensor data over a predefined time window (e.g., 60 seconds), extracts features, and embeds them into a prompt generator. These prompts are then processed by a fine-tuned LLM for interpretation. Finally, the resulting descriptive fragments are aligned and aggregated to generate the user activity logs over a longer period (e.g., one hour).

Spatial Location Awareness: To identify the user's location information, we design a spatial location awareness module to interpret users' location information. By analyzing historical data, we establish a GPS coordinate clustering model for typical locations such as residences, workplaces, and frequently visited places. Wi-Fi Service Set Identifiers (SSIDs) and Bluetooth beacon connection information often contain location-related details, such as "KFC_Free_WiFi" and "$SKY\_Building\_officeRoom1001$". By parsing these character strings and querying global Wi-Fi location databases maintained by companies like Google and Skyhook, the system can further improve the positioning accuracy.

Motion Context Awareness: LLMs exhibit significant limitations in understanding long-duration time-series data. To enhance the reasoning capabilities of LLMs, we explore both time-domain and frequency-domain features of IMU data. We extract a 26-dimensional feature vector to fine-tune the LLM. Next, we analyze audio data to infer contextual information related to user activities, such as attending a lecture in a classroom, conversing at a party, or walking outdoors. However, the inference of raw audio signals has a high computational overhead and contains sensitive user privacy information. To address this, we compute the mean and standard deviation of 13 Mel-Frequency Cepstral Coefficients (MFCCs) along with two parameters related to sound volume intensity, resulting in a 28-dimensional feature vector to fine-tune the LLM. This enables the model to analyze spectral differences across different scenarios and identify more nuanced activity contexts.

Activity Logs Generation: Since environmental and physiological indicators remain relatively stable within a small time window (e.g.,60 seconds), we aggregate these sensor readings by computing their median values and structuring them into a formatted text to enhance inference efficiency. Next, we design a prompt generator to embed the extracted features of multi-modal sensor data into the prompts. As shown in Fig.1, the prompt consists of five key components: data introduction, descriptions of activity features, task explanation, output format, and the specific feature values. Based on the ground truth from dataset, we construct question-answer pairs to fine-tune the LLM and generate textual outputs. Finally, ContextLLM re-input the fine-grained activity log segments into the LLM and merge multiple activity records based on activity type, producing activity log summaries of natural semantic descriptions.

	Precision	Recall	F1-score
Location Recognition	90.72%	91.95%	91.33%
Activity Recognition	88.92%	88.03%	88.47%
Context Recognition	71.23%	70.5%	70.86%
Activity Logs (BERTScore)	79.32%	81.23%	80.26%

Table 2: Preliminary evaluations of ContextLLM

3 EXPERIMENTAL ANALYSIS

We use GPT-4o-mini as the base model to fine-tune ContextLLM and construct a daily life multi-modal sensor dataset based on three real-world datasets. Specifically, location and environmental data come from [6], activity data are sourced from [5], and smartwatch data originate from [8]. Accurate recognition of location, activity, and context is a prerequisite for generating activity logs. Therefore, we first evaluate the performance of ContextLLM on these tasks and then compute the BERTScore to measure the similarity between automatically generated activity logs and human annotations. As shown in Fig.2, clustering based on historical data effectively distinguishes frequently visited locations. Using only GPS data, ContextLLM achieves an F1-score of 91.33% for location recognition. For activity recognition, using raw IMU data as input results in an F1-score of only 21.26%, while our feature extraction and prompt engineering strategies improve it to 88.47%. We compare this result with the recent SOTA approach LLaSA [3], which also utilizes LLMs to interpret this activity dataset [5] and achieves an F1-score of 72%. Our method outperforms it by 16.47%, further demonstrating the advantages of our design. For context recognition, ContextLLM shows a lower performance, with an F1-score of 70.86%, mainly due to the difficulty in distinguishing highly similar scenarios. In future work, we will focus on studying and improving it. The final BERTScore evaluation achieves an F1-score of 80.26%, outperforming AutoLife (70.4%) while it can provide more fine-grained log descriptions.

4 CONCLUSION

We present ContextLLM , a system for generating fine-grained activity logs. Preliminary results show that our carefully designed feature extraction and prompt generation methods are highly effective, significantly improving LLM inference performance and accuracy. The final BERTScore evaluation of ContextLLM achieves an F1-score of 80.26%, demonstrating the feasibility of our approach.

REFERENCES

[1] Adrian Aiordăchioae et al. 2024. Integrating Extended Reality and Neural Headsets for Enhanced Emotional Lifelogging: A Technical Overview. In *2024 International Conference on Development and Application Systems (DAS)*. IEEE, 1–6.
[2] Genki Asano. 2024. LifeLog: Timelog & Diary. https://apps.apple.com/bb/app/lifelog-timelog-diary/id6473384260 2024.
[3] Sheikh Asif Imran et al. 2024. LLaSA: Large Multimodal Agent for Human Activity Analysis Through Wearable Sensors. *arXiv preprint arXiv:2406.14498* (2024).
[4] Ananda Ravuri, Siva S Shankar, DP Devan, Manoj Kumar Padhi, VK Senthil Ragavan, Mahesh Maurya, and A Ravi. 2024. A Systematic Literature Review on Human Activity Recognition. *Journal of Electrical Systems* 20, 6s (2024), 1175–1191.
[5] Jorge-L Reyes-Ortiz, Luca Oneto, Albert Samà, Xavier Parra, and Davide Anguita. 2016. Transition-aware human activity recognition using smartphones. *Neurocomputing* 171 (2016), 754–767.
[6] Yonatan Vaizman et al. 2017. Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE pervasive computing* 16, 4 (2017), 62–74.
[7] Huatao Xu, Panron Tong, Mo Li, and Mani Srivastava. 2024. AutoLife: Automatic Life Journaling with Smartphones and LLMs. *arXiv:2412.15714* (2024).
[8] Sofia Yfantidou et al. 2022. LifeSnaps, a 4-month multi-modal dataset capturing unobtrusive snapshots of our lives in the wild. *Scientific Data* 9, 1 (2022), 663.