深度學習於感測器資料驅動下之人類活動識別研究

學生:許安　指導教授:曾新穆 講座教授

Introduction

Human Activity Recognition (HAR) is a burgeoning field focused on developing systems capable of identifying and categorizing human activities through sensor data analysis. Utilizing accelerometers, gyroscopes, and environmental sensors, HAR systems aim to accurately label behaviors like0 running or ascending stairs. The applications of HAR are vast, ranging from fitness monitoring to healthcare diagnostics and immersive technologies like augmented reality. The sensor-based approach, particularly with wearable devices, offers unobtrusive and privacy-preserving continuous monitoring, providing rich contextual details crucial for distinguishing diverse human activities.

Contributions:

- We introduce and implement cutting-edge neural network architectures, combining Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and advanced attention mechanisms to enhance the recognition and categorization of human activities.
- We achieve exceptional model accuracy, ranging from 93% to 99%, through this integrated approach that leverages the WISDM and OPPORTUNITY datasets, thereby demonstrating the effectiveness of our innovative methods in the precise identification of human activities..
- We conduct an extensive comparative analysis, unveiling crucial insights into performance variations linked to our unique model design and the specificities of the datasets utilized.

Overview

As illustrated in Figure 1, our model seamlessly integrates a series of convolutional layers for extracting essential features, complemented by recurrent layers that discern temporal dependencies. It further incorporates attention layers, focusing on critical data aspects. A fully connected layer follows, adeptly handling the classification task. This strategic combination ensures a deep understanding of the intricate patterns inherent in human activities. The model's architecture is detailed in the subsequent flow diagram.

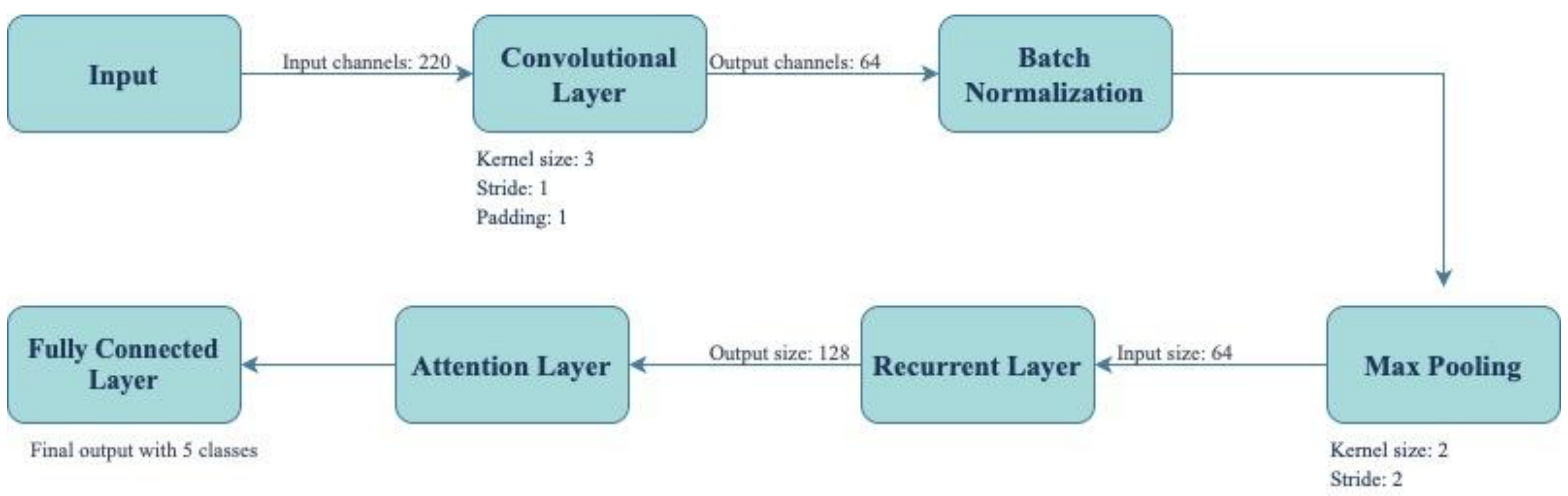

Figure 1. Detailed Architecture of our Neural Network Model.

Material & Methods

Dataset

- *WISDM (Kwapisz et al., 2010)*

  The WISDM dataset was gathered from 29 participants using Android smartphones, records tri-axial accelerometer data for activities including walking, jogging, climbing stairs, sitting, and standing. Supervised by a dedicated researcher, the dataset comprises 1,098,207 samples, segmented into ten-second intervals with features generated from 200 readings.

- *OPPORTUNITY (Roggen et al., 2012)*

  The OPPORTUNITY dataset, collected from 4 subjects using body-worn, object, and ambient sensors, captures daily morning routines in a simulated studio. It provides annotations across various levels, including modes of locomotion, low-level actions, mid-level gestures, and high-level activities. In this research, we focused on 'locomotion' and 'high-level' activities.

Method

- *Data Preprocessing*

  Missing values are interpolated, and standard scaling is applied to normalize the data. Processed data is then segmented into non-overlapping windows.

- *Convolutional Neural Network (CNN) Layer*

  The CNN layer applies 1D convolution with 64 filters, a kernel size of 3, a stride of 1, and padding of 1. The output is normalized, followed by 1D max pooling with a pool size of 2.

- *Recurrent Neural Network (RNN) Layer*

  Long Short-Term Memory (LSTM) is employed with 64 hidden units and the specified number of layers. The output is set to be batch-first for further processing.

- *Attention Layer*

  Implements linear layer to compute attention weights, emphasize specific temporal information in the input sequence, enhancing the model's ability to capture relevant patterns for classification.

- *Fully Connected Layer*

  Implements linear layer to map features extracted from the preceding layers to the output space. This layer produces logits for classification, serving as the final step in the model's architecture.

Results

In our experiments, the same hybrid neural network model was rigorously trained and tested on both the WISDM and OPPORTUNITY datasets. As shown in Table 1, on the WISDM dataset, the model attained metrics including an accuracy of 93.64%, precision of 90.54%, recall of 90.95%, F1-score of 91.89%, and AUROC of 98.35%. In contrast, training with the OPPORTUNITY dataset, the model exhibited superior performance, reflected in the model of high-level activities having an accuracy of 97.42%, precision of 97.42%, recall of 97.42%, F1-score of 97.42%, and an AUROC of 99.78%, and the model of locomotion having an accuracy of 99.65%, precision of 99.60%, recall of 99.60%, F1-score of 99.60%, and an astounding AUROC of 99.99%,

As illustrated in Figure 2, the model exhibits overall better performance on the OPPORTUNITY dataset compared to the WISDM dataset. This enhanced performance can be attributed to the OPPORTUNITY dataset being collected using seven inertial measurement units and twelve 3D acceleration sensors, provides a more informative input space compared to the WISDM dataset, which relies solely on a 3-axis accelerometer. This richer sensory information in the OPPORTUNITY dataset likely contributes to its superior performance.

Despite being collected from a smaller number of participants (only four subjects), the OPPORTUNITY dataset exhibits greater diversity and comprehensiveness in terms of the types of activities and scenarios it covers. This diversity enhances the model's ability to generalize to a broader range of scenarios.

Furthermore, our results indicate that the model performed better on locomotion classification tasks within the OPPORTUNITY dataset. This improved performance can be attributed to the structured nature of locomotion activities, with their well-defined and distinguishable patterns, allows the model to learn and recognize key features more effectively. In contrast, high-level abstractions might introduce greater variability, making it challenging for the model to capture and generalize essential features for accurate recognition.

Table 1. Model Performance Comparison of OPPORTUNITY and WISDM dataset

Dataset / Task	Accuracy	Precision	Recall	F1-scroe	AUROC
WISDM	93.64%	90.54%	90.95%	91.89%	98.35%
OPPORTUNITY (High-level)	97.42%	97.42%	97.42%	97.42%	99.78%
OPPORTUNITY (Locomotion)	99.65%	99.60%	99.60%	99.60%	99.99%

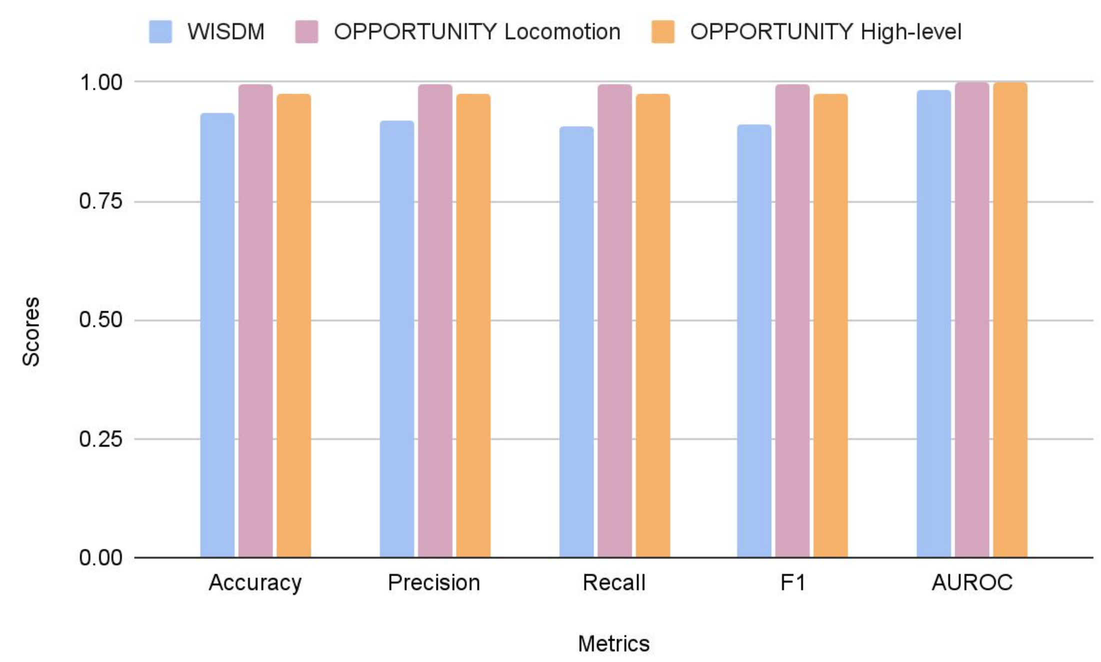

Figure 2. Comparative Analysis of Performance Metrics for the WISDM Dataset and Two Tasks within the OPPORTUNITY Dataset.

Conclusion

In this project, we delved into the realm of Human Activity Recognition (HAR) with a focus on developing and comparing enhanced neural network models for activity classification using the WISDM and OPPORTUNITY datasets. Employing a hybrid model that integrates CNN, RNN attention mechanisms, and fully connected layers, the resulting models demonstrated great results in each performance metrics, with the model trained with the WISDM dataset having an accuracy of 93.64%, precision of 91.89%, recall of 90.54%, F1-score of 90.95%, and AUROC of 98.35%, and with the model trained with the OPPORTUNITY dataset having an accuracy of 99.24%, precision of 99.25%, recall of 99.24%, F1-score of 99.24%, and AUROC of 99.99%. However, a notable discrepancy in performance was observed, with the OPPORTUNITY dataset outperforming the WISDM dataset across all metrics.

This underscores the importance of considering dataset characteristics, such as sensory input richness, diversity, and annotation depth, in the development of robust HAR models. The proposed neural network architecture, combining Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), attention mechanisms, and fully connected layers, proved effective in handling the complexities of activity recognition.

In future work, the exploration of larger and more diverse datasets, along with the incorporation of advanced techniques such as transfer learning and ensemble methods, could further enhance the generalization capabilities of HAR models. Additionally, real-world deployment scenarios and considerations for model interpretability and explainability should be explored to ensure the practical applicability of HAR systems in various domains, including healthcare, fitness, and immersive technologies.