Georgios Kontogiannis 1070908
Ioannis Velgakis 1047071

Gesture Recognition for the identification of common phrases in ASL

AIoT Course – Human Gesture Recognition Project

Algorithmic Foundations of Sensor Networks, CEID, University of Patras

May, 2023

Abstract

End-to-end Artificial Intelligence of Things project about developing a system for recognizing automatically six common phrases in American Sign Language (ASL). Using MetaMotionR, a wristband equipped with two sensors, a gyroscope and an accelerometer, kinesiological data of the gestures were recorded, processed and analyzed in order to train and evaluate three statistical classifiers and two neural network classifiers. The SVM-optimized classifier emerged as the most accurate, followed closely by the RandomForest and CNN classifiers.

Introduction

Figure 1: ASL gestures used for the project

The development of a robust and accurate ASL gesture recognition system can greatly enhance communication and interaction for individuals who rely on sign language. This report presents a procedure and evaluation of an ASL gesture recognition system using MetaMotionR's accelerometer and gyroscope sensing modules. The objective is the identification of six ASL gestures: {"Hello," "Goodbye," "Thank you," "Please," "Yes," and "No"} (Fig. 1). The project follows a systematic procedural pipeline, including data collection, exploratory data analysis / data engineering and learning process in order to identify the optimal classifier and assess its accuracy.

Data Collection

For the data collection, MetaMotionR research sensor kit, and its wristband, which is a wrist-worn device that provides recorded (logging) or real-time (streaming) sensor data was used. These sensors capture the motion and orientation of the wrist, enabling the detection of hand gestures associated with ASL. The

collected ASL gestures/classes were: {"Hello", "Goodbye", "Thank you", "Please", "Yes", and "No"}. Consecutive start and stop times were meticulously documented for each individual instance throughout the execution of the gesture, ensuring minimal interference from extraneous factors. The sampling frequency for both sensors was 100 Hz. Each gesture instance was performed periodically with a fixed reference point for approximately 7 seconds, resulting in around 800 samples per instance. A total of 300 instances were collected, with 50 instances for each class. The collected data were uploaded to a Google Drive directory and stored for further analysis and processing. MongoDB was employed as a database repository for efficient data management.

Exploratory Data Analysis and Data Engineering

The collected data were retrieved from the database and organized into a gesture dataframe, consisting of a list of data instances for each ASL gesture class. The dataset comprised 300 samples, with 50 samples per class. To increase the quantity, improve the robustness of the dataset and avoid overfitting, a sliding window technique was applied. A window size of 150 samples was chosen, resulting in 2667 augmented data samples (2667, 150, 6). Each data sample consisted of six features: acc_x, acc_y, acc_z, gyr_x, gyr_y, and gyr_z. It should be noted that if the remaining samples at the end of the dataset were not sufficient to form a complete window of size 150, the last window was discarded to ensure consistency in the analysis.

The dataset underwent Fourier transform, generating periodograms and spectrograms to visualize the frequency distribution. Notably, distinct peaks were observed at 0 Hz, which corresponded to the gravity component of the sensors, and within the (.4 - .8) Hz range, capturing the dynamic motion characteristics of interest. Moving forward, an optimal order of 1 was determined using the best fit model and Akaike Information Criterion. To isolate the desired frequency range and focus on the dynamic behavior, a band-pass filter was applied at the aforementioned specifications, effectively attenuating unwanted noise and interference while preserving the relevant signals. The preprocessed data samples were further prepared for analysis and modeling. The (2667, 150, 6) tensor was flattened and reshaped into (2667, 900), followed by appropriate scaling. For subsequent statistical learning models, an 80%-20% split was employed to divide the data into training and test sets. For neural network models, a 70%-15%-15% split was utilized for training, test, and validation sets, respectively. To gain insights into the dataset's structure and feature importance, Principal Component Analysis (PCA) was applied. By capturing 90% of the dataset's variance, the optimal number of components was determined to be 12. After dimensionality reduction using PCA, the transformed dataset took on a shape of (2667, 12), facilitating further analysis and modeling.

Learning

Statistical Classifiers

The Support Vector Machine (SVM) and RandomForest statistical classifiers were employed for ASL gesture recognition. For the statistical classifier models each training sample corresponds to a (900,) vector (#columns x window_size) for the flattened – scaled data and (12,) for the variance – PCA data. The classifiers were trained / evaluated for both variations of the data set. For the flattened – scaled data the SVM achieved an accuracy of 0.945 and the RandomForest a score of 0.973 with no optimization. For the latter the SVM achieved an accuracy of 0.932 and the RandomForest an accuracy score of 0.981. Grid

Search and Cross-validation were used to determine the best parameters for the optimized SVM classifier, resulting in an accuracy of 0.990 and a standard deviation of 0.

Neural Network Classifiers

Three neural network classifiers, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, were utilized. Each training input has a shape of (150, 6) - (window_size, #features). The CNN achieved an accuracy of 0.980. The RNN with LSTM architecture achieved an accuracy of 0.170. Similarly, the RNN with GRU architecture achieved an accuracy of 0.115.

Discussion

When analyzing the performance of the models using the input data with 90% variance retained through PCA, we observed a slight increase in false positives for the SVM classifier, resulting in a decrease of approximately 10e-3. This behavior is intuitive because while PCA components are increased, higher separability is easily observed by the PCA 2D / 3D scatter plots. On the other hand, the RandomForest classifier showed positive performance effect since it is less sensitive to the characteristics of the vector space. Choosing to retain 90% of the variance in the PCA analysis proved to be a worthwhile trade-off between accuracy and computational cost, as it reduced the number of components while still capturing a significant portion of the underlying data patterns.

The best training models for ASL gesture recognition using the wristband sensor were the SVM-optimized classifier with an accuracy score of 0.99, followed by the RandomForest classifier with an accuracy of 0.98, and the CNN with an accuracy of 0.98. These results demonstrate the potential of using a wristband-based system for accurate ASL gesture recognition, paving the way for improved communication and accessibility for individuals using sign language.

Conclusion

In conclusion, this project successfully developed an ASL gesture recognition system using a wristband sensor. We achieved high accuracy in recognizing six ASL gestures using the SVM-optimized classifier with accuracy score of 0.99. This result demonstrates that in a larger scale potentially an application for a smart watch could be deployed in order to aid people with hearing impairment problems (smart watches have built-in accelerometer and gyroscope modules). This system has the potential to enhance communication for individuals using sign language and contribute to inclusivity and accessibility in various domains. Further research can be conducted to expand the gesture vocabulary and improve the system's robustness in real-world scenarios.

References

1. “MMR – MetaMotionR,” Mbientlab, [Online]. Available: https://mbientlab.com/store/metamotionr/.
2. Tzamalis, Pantelis, Andreas Bardoutsos, Dimitris Markantonatos, Christoforos Raptopoulos, Sotiris Nikoletseas, Xenophon Aggelides, and Nikos Papadopoulos., “End-to-end Gesture Recognition Framework for the Identification of Allergic Rhinitis Symptoms.,” in 2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS), Marina del Rey, Los Angeles, CA, USA, 2022.
3. Tzamalis, Pantelis., "Python Data Science and Machine Learning Tutorials", [Online]. Available: https://github.com/tzamalisp/data-science-and-machine-learning-tutorials
4. “CS231n Convolutional Neural Networks for Visual Recognition,” cs231n.github.io. https://cs231n.github.io/convolutional-networks/#conv
5. Startasl.com, 2023. https://www.startasl.com/wp-content/uploads/asl-signs-new-705x494.jpeg

Appendix

Google Drive directory: https://drive.google.com/drive/folders/1SbffJaO2COmObzk3Z2Do_USM-ZuBp2KR?usp=sharing