BlueSky Wanderer: Improving Sensor Data Classification Through Real-time Visualization

Marco Cavallo and Çağatay Demiralp

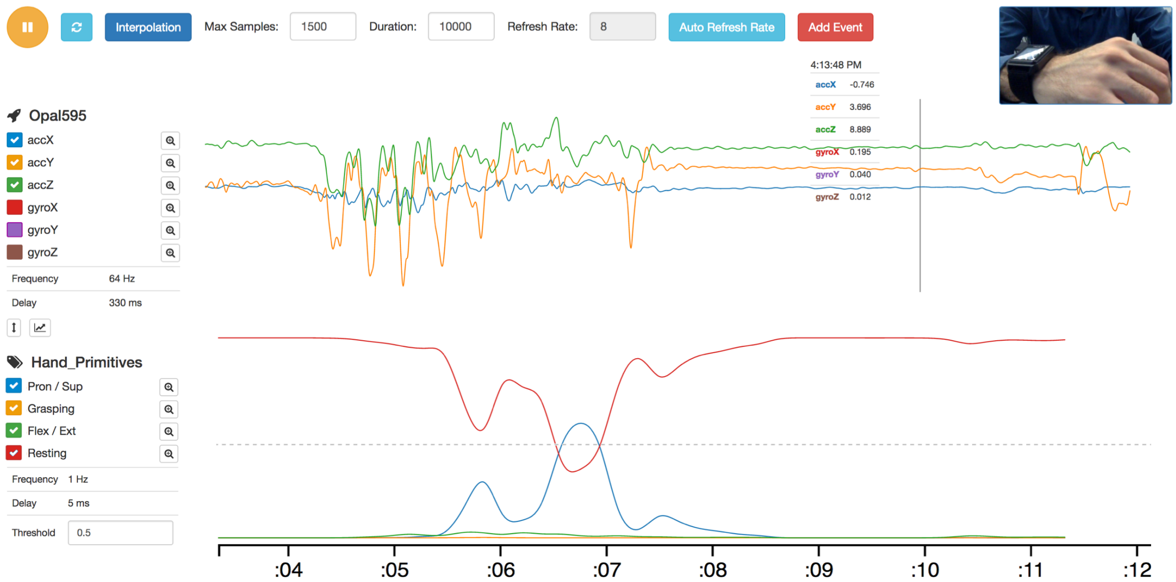

Fig. 1. BlueSky Wanderer user interface. In the upper chart, live sensor data is represented as a line plot, whereas the lower chart shows the activation function of a multi-class classifier. The system enables users to observe in real-time the performance of classification models and to interactively refine them. By brushing on a region of the upper chart, insights derived from statistical and spectral analysis of the time-series data are presented to the user.

Abstract

Sensors ranging from wearables to smart phones are becoming increasingly ubiquitous and essential in our everyday lives. Detecting events or activity in temporal data collected by sensors is an important task in sensor applications.

Temporal data classification often involves training a model on a set of labeled data. While collecting data samples for training is a relatively inexpensive task, obtaining interpretable ground-truth labels can be, however, challenging. Datasets of wearable sensor data often contain limited information about the usage context of the devices and video validation is not always possible. This generally makes it harder for data scientist to interpret the sensor data, sometimes leading to overly-expensive machine learning approaches running as a black-box on massive datasets. It could be beneficial, instead, to first understand the sensors' behavior and then iteratively develop a classification model.

We introduce *BlueSky Wanderer*, an interactive visual analysis tool for running classification algorithms on live-streaming sensor data. Our system, meant to complement offline model training, provides visual and quantitative solutions to iteratively reason on classification performance and to develop better intuition about patterns in data. In particular, *BlueSky Wanderer* offers a set of interactions and summary visualizations to support quick debugging, dynamic fine-tuning and validation of classification models.

- *Marco Cavallo. E-mail: mcavall@us.ibm.com*
- *Çağatay Demiralp. E-mail: cagatay.demiralp@us.ibm.com.*

In our proposed approach, the data scientist wears the sensor devices himself and observes the real-time response of the classification algorithms. While raw sensor and classifier output is visualized as line or block charts, statistical information and insights on the data are computed and presented to the user. Principal component analysis (PCA) and spectral analysis respectively show relevant input signals and dominant frequencies over a selected time period. Model parameters such as threshold and time classification window are visually represented and can be dynamically modified from our tool. A built-in Python code editor further allows the data scientist to refine the model in real-time, observing its change in performance with respect to live-generated motor events. Additional features offer the possibility to annotate events, to save sensor data for further offline analysis, and to export a statistical report for each detected activity.

Here, we use a case to demonstrate how our approach can improve the development of a set of motion classifiers in the context of monitoring the symptoms of Parkinson's disease. In particular, we asked a team of four data scientists to develop models that can recognize four different body movements (hand pronation-supination, hand flexion-extension, body turn, hand tremor), leveraging two IMU sensors placed on the right wrist and on the chest of a subject. The data scientists alternated, two at a time, the use of our visualization system to their usual development workflow. Our results show that, in all cases, *BlueSky Wanderer* proved to be a useful companion in the process of developing classification algorithms on time-series data. All models produced with the help of our system outperformed their counterparts when handling real-life test data. Further, data scientists reported that the visualization system helped them identify much faster relevant patterns in the sensor data.