Event Detection in Sports-Related Sensor Data with Machine Learning

Mallory Cashman
Advisor: Dr. Laura Dietz
Department of Computer Science, University of New Hampshire

Introduction

SPAITR is a UNH student startup with the goal of empowering players to improve their performance by making individual athlete's analytics more accessible. This allows them to make data driven training decisions.

Machine learning models are more accurate and are preferred to the heuristic shot detection method currently in place. This is due to their independence from assumptions about the capabilities of the player and sensors.

Data

- SPAITR's Neuro™ collects time-series accelerometer and gyroscope data in three dimensions
  - μs, ax, ay, az, gx, gy, gz
- Dataset includes a variety of lacrosse activities including:
  - Isolated shots
  - Several shots in sequence
  - Cradling
  - Ground balls
  - Ball drops
- Dataset composition: 145 shots and 145 "not shots"
- Time series split into windows with 70 observations (about 2.8 seconds)
- Goal is to predict if the window overlaps with a shot event

Table 1: Example of shot data including time in microseconds (μs), tangential acceleration (ax, ay, az) in m/s^2, and rotational acceleration (gx, gy, gz) in °/s.

ms	ax	ay	az	gx	gy	gz
0	12960	-10620	5668	315	-1310	-31
40000	13048	-10716	5352	21	-971	-113
80000	13100	-10832	5888	-182	-1185	-214
120000	12892	-10820	5384	-360	-1720	-241
160000	13084	-10980	5792	-356	-406	-172

Goal

Task: Train machine learning models to predict the number of lacrosse shots collected by SPAITR's Neuro™ during a practice session, using timeseries based accelerometer and gyroscope data.

Shot Data Visualization

Figure 1: Phases of a lacrosse shot.

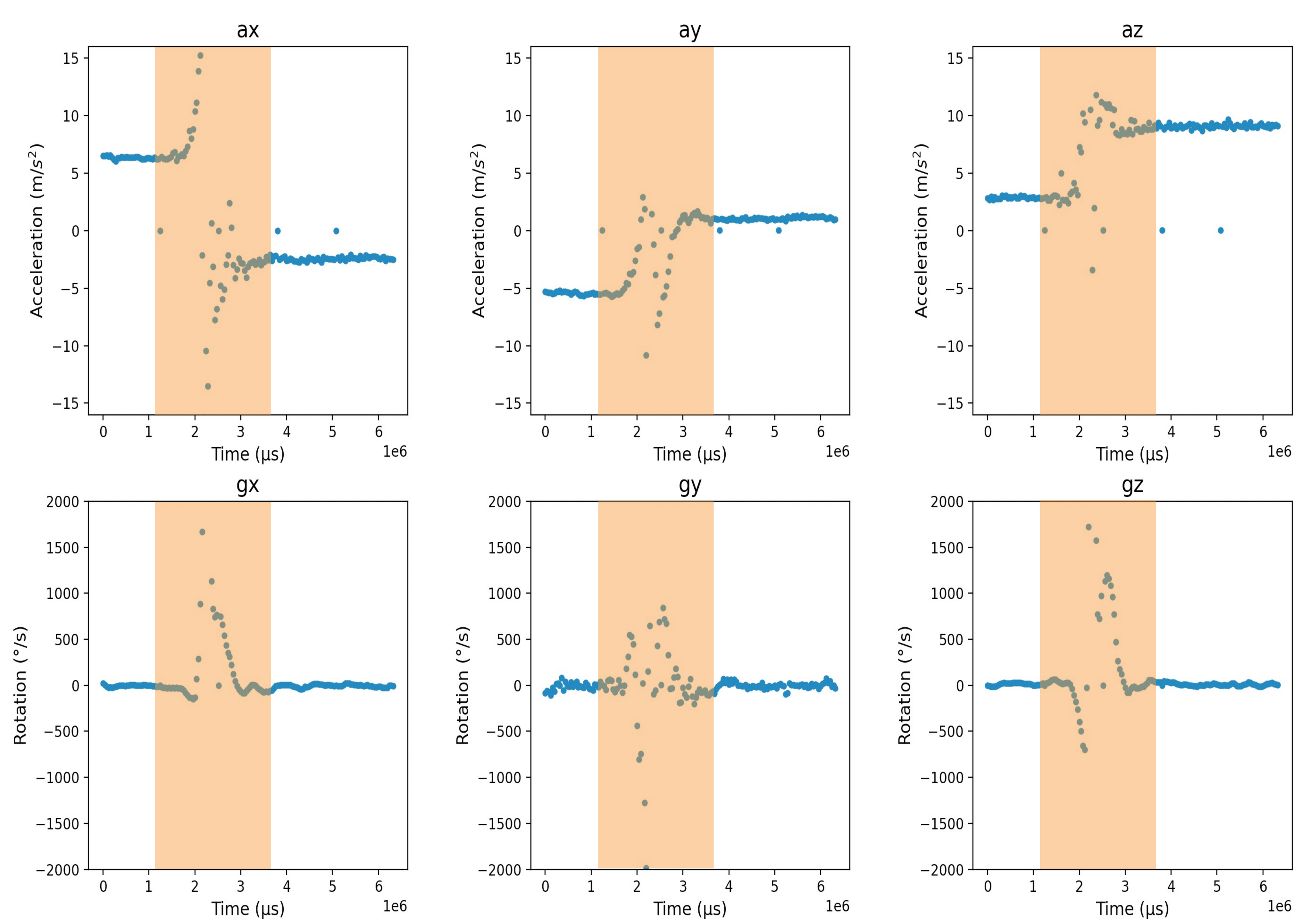

Figure 2: Graphical representation of shot data including the six motion variables over time, with the orange window highlighting the shot itself.

Methodology

- Randomly sampled training and test sets with 70:30 split
- Features are the six variables over all 70 samples in the window (flattened into a vector)
- Our approach predicts the number of shots in the collection period by (1) predicting shots in sliding windows, then (2) collapsing subsequent shot windows
- The scikit learn Python library allows for rapidly developed and adaptable machine learning models
- Baseline models include always positive, always negative, and random classification
- Machine learning methods:
  - Decision Tree

$$Q_m^{left}(\theta) = \{(x, y) | x_j <= t_m\}$$
$$Q_m^{right}(\theta) = Q_m \setminus Q_m^{left}(\theta)$$
$$\theta^* = \operatorname{argmin}_\theta G(Q_m, \theta)$$
$$G(Q_m, \theta) = \frac{N_m^{left}}{N_m} H(Q_m^{left}(\theta)) + \frac{N_m^{right}}{N_m} H(Q_m^{right}(\theta))$$

  - Support Vector Machine

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i$$
$$\text{subject to } y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i,$$
$$\zeta_i \geq 0, i = 1, \ldots, n$$

Results

High accuracy achieved by machine learning on clean data.

Table 2: Quality of predicting sliding windows containing shots.

Method	Precision	Recall
Always Positive	0.57	1
Always Negative	Undefined	0
Random Classifier	0.65	0.66
Decision Tree	1.0	1.0
SVM	1.0	1.0

Conclusions & Next Steps

- SPAITR should implement machine learning models in their shot detection software
- Next steps:
  - Train models to detect other events
  - Generate and train on data from a diverse pool of athletes in terms of age, gender, and ability

Contact Information: malcashman@gmail.com SPAITR.com