# Representation learning from geospatial data:

# **Towards a robust framework for transportation mode detection**

Antoine Nzeyimana, Henry Hui Luan

Department of Geography, University of Oregon

## Abstract

Representation models of spatial data influence the kinds of analyses we can perform and thus affects the conclusions we can make about the data. Choosing between different representation models may also lead to different accuracies in our analysis results. Data representation also impacts the computational performance of our analysis algorithms, an aspect which is critical for the processing big spatial data. Representation learning aims at learning data representations that are most suitable for predictive modeling tasks. We aim to understand the impact of spatial data representation in geospatial problem solving. We want to understand for instance which information is lost, retained or enhanced when we choose to work with one representation over the other. **Here we present a case study of using spectral representation of accelerometer data for transportation mode detection.**

## Contact Information

- Web: https://blogs.uoregon.edu/s3clab/
- Email: anzeyima@uoregon.edu

## Case study: Transportation mode detection (TMD)

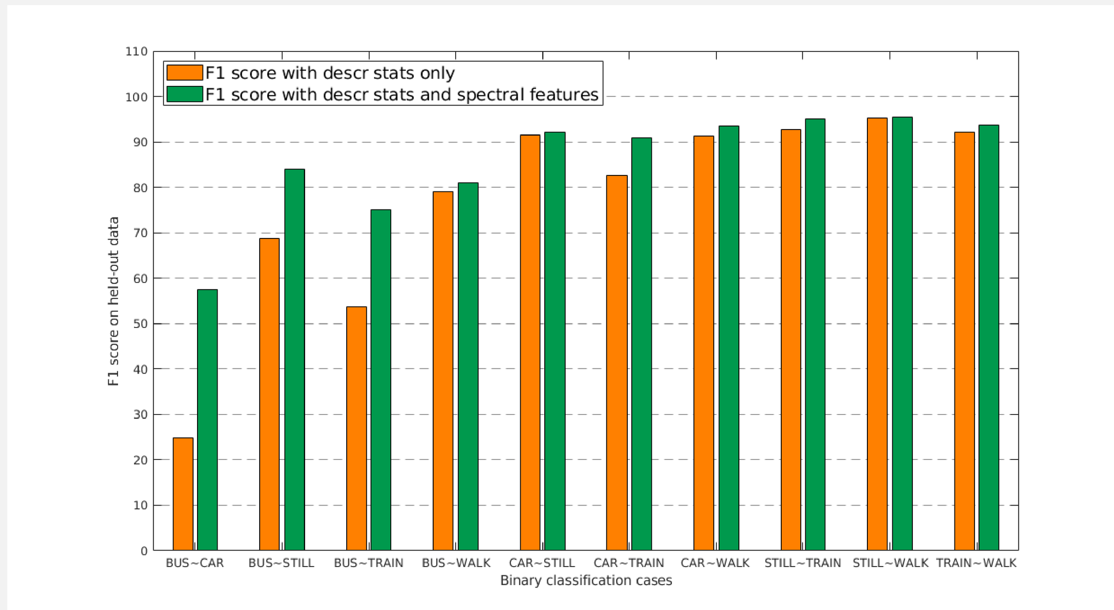
- **Important applications**: Transportation engineering, Public health and population physical activity research, and Context-aware computing.
- **Opportunity**: Ubiquity of different sensing devices (e.g. smart-phones, smart-watches, wearable sensors) providing different sensory information (e.g. GPS tracks, acceleration, speed, sound).

While recent studies have reported high accuracy in TMD, the results are hard to compare due to multiple factors: types and accuracies of the sensory information, sensitivity to sampling rate and data collection methods (e.g. lab settings vs free living conditions). In our study, we assess the impact of input feature representation on the performance of TMD. We report results from an experiment conducted on a publicly available TMD dataset [1] using a standard SVM classifier on the minimum sensory information: non-uniformly sampled raw accelerometer data.

## Methods

- **Features used in a previous study**: Descriptive statistics of a detection window (mean, minimum, maximum, standard deviation, percentiles) of acceleration magnitude.
- **Detrended fluctuation analysis (DFA) [2]**: asseses statistical self-similarity of time-series.
- **Spectral representation via Lomb-Scargle periodogram [3, 4]**: a least-squares spectral analysis of non-uniformly sampled time-series.

## Preliminary results




**Significant improvement in F1 score on hard cases (imbalanced classes).**

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

## Conclusion

We empirically found that fluctuation analysis and spectral representation via Lomb-Scargle periodogram improve transportation mode detection from raw accelerometer time-series. By relaxing the sampling rate uniformity, we hope that this method will accomodate a broad range of devices including low-powered ones. We plan to compare next this method to interpolation and uniform resampling-based power spectrum estimation techniques. We are also going to investigate how to optimally delimitate detection windows based on natural breaks in the data rather than working with arbitrary window sizes.

## References

[1] e. a. Carpineti, Claudia, "Custom dual transportation mode detection by smartphone devices exploiting sensor diversity," in *2018 IEEE PerCom Workshops*. IEEE, 2018, pp. 367–372.

[2] e. a. Peng, C-K., "Mosaic organization of dna nucleotides," *Physical review e*, vol. 49, no. 2, p. 1685, 1994.

[3] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and space science*, vol. 39, no. 2, pp. 447–462, 1976.

[4] J. D. Scargle, "Statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, pp. 835–853, 1982.