

An Improved Data Representation for Smoking Detection With Wearable Respiration Sensors

Laura Hiatt

*Mathematics and Computer Science
Davidson College
lahiatt@davidson.edu*

Roy J. Adams and Benjamin M. Marlin

*College of Information and Computer Sciences
University of Massachusetts Amherst
{rjadams,marlin}@cs.umass.edu*

1. Introduction

According to the CDC, tobacco use is responsible for over 480,000 deaths and \$300 billion in healthcare spending per year. ¹ Within the field of mobile health, cigarette smoking detection using wearable sensors is a key problem with the potential to improve health outcomes by enabling continuous monitoring and personalized, adaptive cessation interventions [1, 2].

Prior mHealth research on smoking detection uses data from a wireless respiratory inductance plethysmography (RIP) chestband sensor alone [1] or in combination with an actigraphy wristband sensor [2] to classify individual respiration cycles as smoking puffs or non-puffs. To solve this problem, the RIP sensor data are pre-segmented into respiration cycles, and feature vectors are extracted from each segment. However, the noise in the RIP signals that results from collecting data in ambulatory settings makes accurately segmenting respiration cycles challenging (see Figure 1). In this work, we propose a data representation for mobile respiration-based smoking detection that is more robust to noise.

2. Approach

We propose a peak-based representation for smoking detection from RIP data. The rationale for preferring this approach in the presence of noise is that there is much less uncertainty about the location of the peaks in the RIP data than there is about the location of segmentation boundaries.

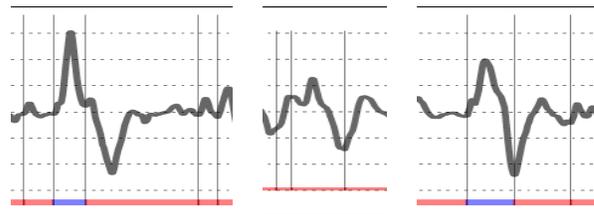


Figure 1: Examples of RIP segmentation errors. Vertical lines indicate segment boundaries. In the data set of Ali et al. [1], we see multiple instances of exhalation being split from inhalation due to noise.

To implement this approach, we first run a peak detection algorithm on the RIP waveform data to extract the peaks (maximum inhalation) and valleys (maximum exhalation) from the input time series. We then extract feature vectors analogous to those used in the segmentation case [1], but defined using relative locations and amplitudes of peaks and valleys. We associate puff (positive) and non-puff (negative) labels with peaks, and learn a classification model to predict the peak labels given the associated feature vectors.

3. Experiments and Results

We compare our approach to that of Ali et al. using the data set described in [1]. We apply a standard leave-one-subject-out evaluation protocol with nested 8-fold cross validation to select classifier hyper-parameters. Similar to [1], we explicitly re-balance the data set to achieve a 50/50 class balance with 328 total examples. We convert the original labeled segments into labeled

¹http://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts/

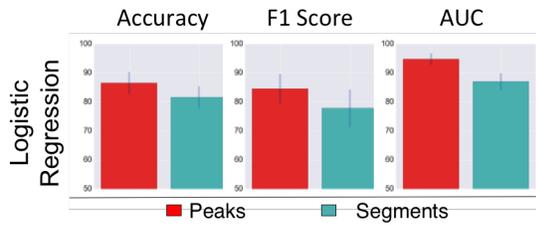


Figure 2: Performance metrics for peaks vs segments using logistic regression. The proposed approach yields improvements for all metrics.

peaks by assigning the highest peak in each smoking segment a positive label. All other peaks receive a negative label. To compare representations, we project our predicted peak labels onto the original segments such that a segment is labeled positive if it contains a peak labeled positive. We evaluate logistic regression (LR) and random forest (RF) classifiers under both representations in terms of accuracy, F1 score, and AUC. The peak representation results in improvements of 5 to 8% for all

metrics for both classifiers. Results for LR are shown below.

Acknowledgments

This work was partially supported by the National Science Foundation under awards IIS-1350522 and CCF-1461021.

References

- [1] A. A. Ali, S. M. Hossain, K. Hovsepian, M. M. Rahman, K. Plarre, and S. Kumar, “mPuff: Automated Detection of Cigarette Smoking Puffs From Respiration Measurements,” in *IPSN*. ACM, 2012, pp. 269–280.
- [2] N. Saleheen, A. A. Ali, S. M. Hossain, H. Sarker, S. Chatterjee, B. Marlin, E. Ertin, M. Al’Absi, and S. Kumar, “puffMarker: A Multi-sensor Approach for Pinpointing the Timing of First Lapse in Smoking Cessation,” in *UBI-COMP*, 2015, pp. 999–1010.