Automated Detection of Anaerobic and Ventilatory Thresholds From Biometric Data

Whitepaper by OMsignal
(Dated: December 12, 2017)

Abstract*—A system is presented for the automated detection of the heart rate that corresponds to a persons ventilatory and anaerobic thresholds (VT and AT). The system is based on an analysis of heart rate, breathing depth and breathing rate measurements from OMsignal apparel while users run. In most cases, the system is capable of providing a reasonable estimate of the AT and VT after 5 free form runs, ie. without a pre-specified protocol. The system automatically determines the AT and VT from this data, closely approximating the values determined by the previously used manual process relying on a human annotator. Due to the difficulty of analyzing free form running data, a sufficiently accurate algorithm has thus far been elusive. Using a sequence of filters and algorithms, automatic AT and VT determination is cast as a computer vision problem, and finally solved with a random forest. On a test set of 40 users, the AT and VT determined by the system were both within 10 BPM of the manually determined values in 95% of the cases, and were both within 7.5 BPM in 85% of the cases. The largest differences occurred on users that had ventilation-HR curves that showed ambiguous inflection points or low variation. Additionally, the system's AT and VT assessments were compared against ground truth values determined by a metabolic cart system on 10 users while they underwent an incremental exercise protocol. In all but one case, the difference was less than 7.5 BPM, and the mean absolute error of VT and AT were 3.56 BPM and 4.75 BPM respectively. The remaining errors are likely due to inherent uncertainty of using free form data as opposed to a regimented protocol. Use cases requiring more accuracy or precision should use such a protocol.

I. Introduction

A system is presented for the automated detection of the heart rate that corresponds to a person's ventilatory (VT) and anaerobic (AT) thresholds from five session of free form running data. Normally, this requires a metabolic cart to measure O_2 and CO_2, along with a regimented incremental exercise protocol. Figure 1 demonstrates the gold standard method for determining AT and VT. Previous work has shown that the breathing measure on OMsignals apparel is sufficient to be used for this purpose: Verification Report: Breathing Depth Algorithm Phase 2. Previous work has also shown that a reasonably accurate assessment of AT and VT values can be determined manually from free form data: VT / AT Validation The purpose of this work is to demonstrate that this can be determined automatically after 5 free form runs.

Figure 2 shows the overview of the system, which is described in detail in sections II through VI.

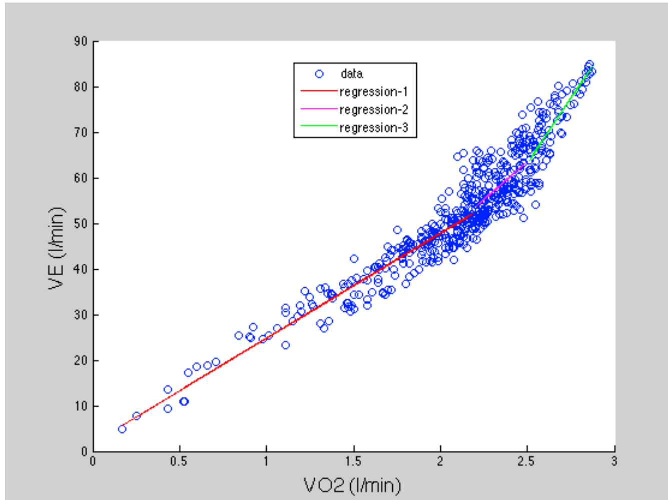

Fig. 1: An example of the gold standard method of obtaining VT and AT by plotting ventilation to oxygen consumption for an incremental exercise test. VT and AT are found by plotting ventilation against VO_2 and then looking for inflection points on the curve as the exercise intensity increases in an incremental exercise test. The inflection at the intersection of regression-1 and regression-2 is VT and the infelction at the intersection of regression-2 and regression-3 is AT. AT and VT values are commonly translated to the heart rate at which the points occurred, so that athletes can use a commercial heart rate monitor to track their zones outside of the laboratory environment.

II. Data Cleaning and Segmentation

All available accounts with manually annotated AT and VT values were assessed for suitability. This involved looking at the ventilation-HR curve of the first five runs and the manually annotated AT and VT values to determine whether there is a suitable variability in heart rate, and that there are no obvious annotation errors. Some accounts were also rejected due to less than 5 minutes worth of data remaining after the biometric filter, described in section 3. After visual assessment, 193 accounts were suitable for inclusion in the algorithm development and validation. These 193 accounts were randomly divided into:

- A training set of 120 accounts

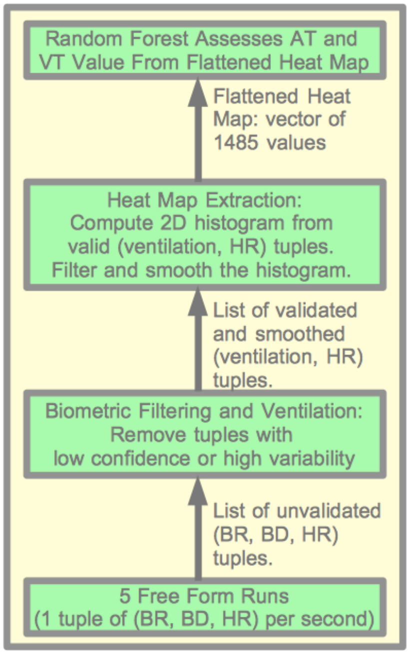

Fig. 2: Flowchart for assessing an AT and a VT from the system. The input is Breathing Rate (BR), Breathing Depth (BD), and Heart Rate (HR), as well as the number of RR intervals, for every second of each of the five runs. Tuples with low confidence or high variability are removed by the biometric filter, breathing variables are smoothed, and the ventilation is computed. The heat map is computed, which then provides a fixed size representation to the random forest model.

- A validation set of 33 accounts
- A final test set of 40 accounts

III. Biometric Filtering

- RR intervals, inhale amplitudes, and inhale to inhale intervals found by OMsignal's biometric algorithms are processed to give one tuple of breathing rate (BR, in breaths per minute), breathing depth (BD, which has no unit), and heart rate (HR, in BPM) per second during each of the five runs.
- These are processed though a filter that removes any second with:
  - A BR above 90, or below 10.
  - A BD of 127 or 0 (which are error conditions).
  - An HR above 200, or below 110 (not useful for AT and VT assessment).
  - An estimated RR coverage within a 50 second window prior to the tuple under consideration below 80%.
  - A standard deviation of HR in a 50 second window prior to the tuple under consideration below 10, to ensure steady state HR.
  - A standard deviation of BR in a 50 second window prior to the tuple under consideration below 15, to ensure steady state BR..

The output of the biometric filter is a sequence of (ventilation,HR) tuples for each of the five runs, called the valid tuples.

IV. Data Augmentation

The data is then augmented through a process designed to partially correct for the small size of the dataset, as well as to remove the bias of the existing dataset. The augmentation process, applied to each users data is as follows:

- Multiple copies of the valid (ventilation,HR) tuples from the first five runs are created by repeatedly drawing bootstrap re-samples of the set of tuples. This introduces variability in the density of the ventilation-HR curve while preserving the general shape and any inflection points.
- Multiple copies of each re-sampled group of five runs are created by shifting the HR values based on a target VT between 120 BPM and 175 BPM inclusive by adding $HR_{shift} = VT - VT_{target}$. The target AT is similarly determined by adding HR_{shift}. Only copies with $AT_{target} \leq 195$ are included.

V. Data Preprocessing and Heat Map Extraction

All five runs are concatenated together. The distribution of these points are approximated with a 2D histogram, to which further processing is applied. The resulting representation is called the heat map. Examples are given in Figures 3, 4 and 5

- All tuples with ventilation values in the 1st and 99th percentiles are removed. This is done because these points tend to be outliers which can significantly degrade the quality of the heat map.
- Ventilation is normalized to lie between 0 and 1, since it is a unit-less quantity in the OMsignal system and its absolute value is not meaningful for AT and VT assessment.
- The 2D histogram has 45 bins for HR spanning [110,200], and 33 bins for ventilation.
- The histogram is filtered with a 3x3 moving maximum filter followed by a 3x3 moving average filter.
- All bins below the maximum entropy density are set to zero.
- All bins above the 95th percentile of the histogram bin densities are set to that value.
- The histogram is re-normalized to sum to 1.

Steps four to six are applied to correct for the fact that the density tends to be highly concentrated in a few bins. Since this is free form running data, the users tend to spend a lot of time at only a handful of HR values.

VI. Model Structure and Training Procedure

A flattened heat map is the result of the data pre-processing stage. From this a random forest is used to simultaneously predict an AT and a VT value for the superimposition heat map. The random forest is trained with the following hyperparameters, chosen by trial and error based on the validation data performance:

- Number of Trees: 3000
- Loss Function: mean squared error
- Minimum Samples per Leaf: 100
- Number of Features per Tree: 39

Of these, the Number of Features per Tree was found to have the greatest effect on generalization error.

VII. Results

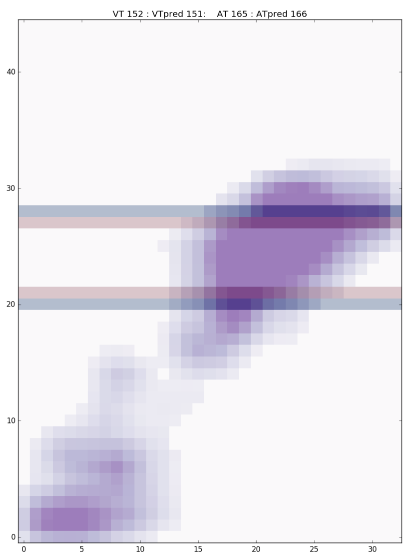

Fig. 3: Example heat map. Ventilation along x-axis, heart rate along y-axis. Demonstrates good inflection points obtained from free form run data. Manual (red bars) and algorithmic (blue bars) AT and VT values are very close.

The performance results are presented in two phases. The first phase model is trained on the training set with hyper-parameters chosen based on validation set performance. The phase two model is trained on both the training and validation sets with the same hyper-parameters as the phase 1 model. Table I shows the performance of the phase 1 model on the augmented datasets. Table II shows the results of the phase 2 model on the augmented datasets. Table III shows the results of the phase 2 model on the un-augmented datasets.

IV compares the AT and VT values determined by the algorithm from free form data to the gold standard values determined from the metabolic cart in an incremental exercise protocol.

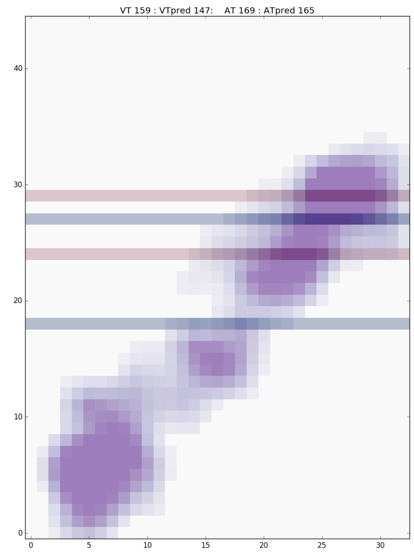

Fig. 4: Example heat map. Ventilation along x-axis, heart rate along y-axis. Demonstrates one form of ambiguity that can occur: there is not clear VT inflection point. As a result the manual VT (lower red bar) and algorithmic VT (lower blue bar) are far apart.

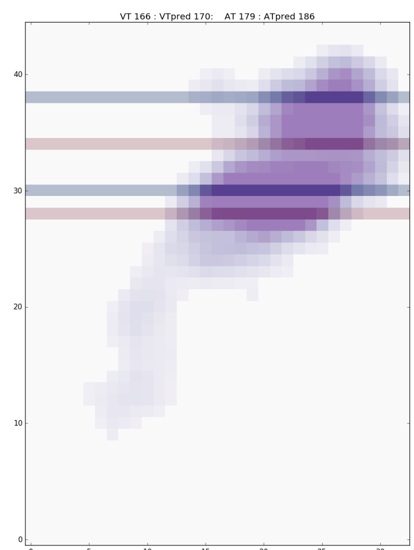

Fig. 5: Example heat map. Ventilation along x-axis, heart rate along y-axis. Demonstrates a thick cloud, it is not visibly obvious where in the cloud the inflection points are. Note that the visible inflection point in the bottom half is not AT or VT, but corresponds to an inflection point typically seen at the transition from walking to running.

VIII. Discussion and Conclusion

In addition to a random forest model, several neural network architectures were also considered. These attempted to analyze each one of the runs individually, and then merge the analysis together. In all cases, however, the neural networks performed comparably or worse on the validation set than the random forest, while taking significantly longer to train. It is hypothesized that this is due to the fact that the current error level is already very near what is possible from free form run data. If true, this would imply that any model with sufficient capacity would be able to generalize with a similar error rate, regardless of its inductive bias. Evidence for this is provided by the relatively small reduction in error afforded by the extra training data available to the model in phase 2 versus phase 1.

An investigation into the errors on the test set suggests that improvement is not possible by modifying the machine learning model. Figure 3 shows a heat map representation with a well formed ventilation-HR curve from the test set. On this example, the manual and algorithmic AT and VT values are very close. Figure 4 shows the heat map for one of the largest VT errors in the test set. Given the ambiguous nature of the curve, it is unlikely that an algorithm could be reliably more accurate in this kind of situation. A human annotator would have difficulty in this situation as well due to the lack of obvious inflection points.

An acceptable level of error for the AT and VT was chosen to be 7.5 BPM. 85% of the test data falls within this acceptance threshold. A manual investigation of the test data comparing the manual and algorithmic AT and VT values revealed that the remaining differences are likely due to inherent uncertainty stemming from the use of wild data. In particular, this data can present thick clouds as in Figure 5 and ambiguous or no apparent inflection points as in Figure 4. It is worth noting that two human annotators may differ by more than 7.5 BPM a small percentage of the time. Given the high correspondence between the gold standard AT and VT values and the algorithmically assessed values (only user 9 had an AT outside of tolerance), this is an acceptable level of error.

Incremental exercise protocols with OMsignal apparel typically result in far less ambiguous ventilation-HR curves. Use cases that require a greater accuracy or precision than is achieved here should consider using such protocols in order to minimize the uncertainty introduced by free form activity.

TABLE I: Performance of the phase 1 model on the augmented data. All errors are absolute errors. The 95% error columns indicates the 95th percentile of the absolute error. Jointly within x BPM indicates the condition where both the AT error is $\leq x$ and the VT error $\leq x$.

Dataset	VT 95% Error (BPM)	VT Worst Error (BPM)	AT 95% Error (BPM)	AT Worst Error (BPM)	% Jointly Within 5 BPM	% Jointly Within 7.5 BPM	% Jointly Within 10 BPM
Training	4.95	11.1	4.85	14.78	91.97%	99.32%	99.98%
Validation	9.05	20.78	8.4	12.69	53.54%	80.19%	94.71%
Test	9.51	14.24	7.52	16.37	59.94%	84.56%	95.71%

TABLE II: Performance of the phase 2 model on the augmented data. All errors are absolute errors. The 95% error columns indicates the 95th percentile of the absolute error. Jointly within x BPM indicates the condition where both the AT error is $\leq x$ and the VT error $\leq x$.

Dataset	VT 95% Error (BPM)	VT Worst Error (BPM)	AT 95% Error (BPM)	AT Worst Error (BPM)	% Jointly Within 5 BPM	% Jointly Within 7.5 BPM	% Jointly Within 10 BPM
Training + Validation	4.02	13.84	3.44	11.73	96.67%	99.43%	99.95%
Test	9.03	15.21	7.61	17.12	58.86%	84.63%	96.37%

TABLE III: Performance of the phase 2 model on the un-augmented data. The performance on the test set is similar to the performance on the augmented data. All errors are absolute errors. The 95% error columns indicates the 95th percentile of the absolute error. Jointly within x BPM indicates the condition where both the AT error is $\leq x$ and the VT error $\leq x$.

Dataset	VT 95% Error (BPM)	VT Worst Error (BPM)	AT 95% Error (BPM)	AT Worst Error (BPM)	% Jointly Within 5 BPM	% Jointly Within 7.5 BPM	% Jointly Within 10 BPM
Training + Validation (N=153)	3.14	14.46	2.77	8.85	97.39%	98.69%	99.35%
Test (N=40)	8.61	14.65	7.19	8.64	60.00%	85.00%	95.00%

TABLE IV: Comparison between algorithmically determined AT and VT (columns labeled Predicted) from free form data, and AT and VT values from a metabolic cart and an incremental exercise test (columns labeled True). The final row shows the mean absolute error for AT and VT. All values are in BPM. The algorithm was not able to make a prediction for user 1 due to very poor quality data (the biometric filter removed almost all of this users data). * indicates the person was included in the training set.

User ID	True VT	True AT	Predicted VT	Predicted AT	Absolute VT Error	Absolute AT Error
1	136	156	N/A	N/A	N/A	N/A
2*	159	184	156	180	3	4
3*	158	N/A	155	172	3	N/A
4	141	168	144	161	3	7
5	165	182	162	175	3	7
6*	157	178	152	175	5	3
7	152	167	146	163	6	4
8	155	167	152	170	3	3
9	155	179	152	170	3	9
10	158	173	161	174	3	1
Mean					3.56	4.75