Proceedings of the IEEE SoutheastCon 2015, April 9 - 12, 2015 - Fort Lauderdale, Florida

Detection of Cigarette Smoke Inhalations from Respiratory Signals Using Decision Tree Ensembles

Daniel Berry
Department of Information Systems,
Statistics, and Management Science
University of Alabama
Tuscaloosa, Alabama 35487
Email: dberry@crimson.ua.edu

Jason Bell and Edward Sazonov
Department of Electrical and
Computer Engineering,
University of Alabama,
Tuscaloosa, Alabama 35487

***Abstract*—In this study we explored the ability of ensembles of decision trees to classify hand-to-mouth gestures in order to detect cigarette smoke inhalations. Three subject independent models were constructed using a variety of ensemble techniques: boosting (AdaBoost), bootstrap aggregating (bagging), and Random Forests. Data was gathered during previous studies by extracting features from the signal waveforms of worn sensors. Each hand gesture was associated with either a smoke inhalation or a hand gesture of another type (e.g. eating). Subject as well as group models were trained. For the group models, model performance was evaluated by computing F-score, precision, and recall statistics using a 20-fold leave-one-out cross-validation testing strategy where one subject was held out for evaluation and models were trained on the remaining 19 subjects. For the individual models, models were trained on a single subject and evaluated using 5-fold cross validation. The average and standard deviation of each statistic across all folds were reported.**

Keywords– Smoking, wearable sensors, decision tree ensembles, boosting, Random Forest, inter- and intra-subject variability

I. Introduction

Tobacco, particularly cigarette smoking, is a leading cause of preventable death in the United States . Cigarette smoking is a known major risk factor for the top 4 leading overall causes of death: diseases of the heart, cancer, chronic lower respiratory diseases, and cerebrovascular disease (stroke) [1]. Significant efforts have been undertaken to study smoking habits. The simplest form of which is to ask patients to self-report cigarette intake. It is well known that patients tend to under report their levels of tobacco consumption, potentially by up to 25% or more [2]. The clinical importance of accurate, noninvasive smoking monitoring systems to quantify intake cannot be overstated. An ideal system would be able to quantify clinically important features such as quantity of cigarettes consumed, volume of smoke inhaled, etc. Previous efforts have included devices to monitor patients' smoking topography. Smoking topography is the quantification of cigarette puff characteristics such as puff duration or puff velocity [3]. Unfortunately, systems such as these are not completely noninvasive and do not fully satisfy the requirements of a free-living system.

Personal Automatic Cigarette Tracker (PACT) systems were designed to quantify smoking behavior in free-living environments [4], [5]. The system uses a wearable Respiratory Inductive Plethysmograph (RIP) and proximity sensor (PS) (arranged in a hand-to-mouth configuration) to record breathing patterns and hand motions under smoking and non-smoking conditions. Previous work on these systems includes the development of a Support Vector Machine (SVM) classifier using the raw sensor signals. More recently, efforts were undertaken to extract interpretable features from the raw sensor signals. In that study, 27 empirically defined features were computed from the sensor waveforms of hand-to-mouth gestures and submitted to a SVM classifier.

II. Methodology

A. Sensor system

In this section we briefly describe the sensor system used in the experiments. This system is more thoroughly described in [6], [7]. A commercially available Respiratory Inductive Plethysmograph (RIP) (Pro-Tech Inc.) was used to record the subjects' breathing. This consisted of two elastic monitoring bands: a thoracic band and an abdominal band (DuraBelt). These sensors quantified the change in the volume of subjects' chests during both smoking and non-smoking activities. Output signals were labeled $TC(t)$ and $AB(t)$ from the thoracic and abdominal bands respectively. Hand-to-mount gestures were gathered using a transmitter-receiver proximity sensor system.

B. Data Collection

Experiments were conducted using 20 regular, long-term (>1 year) smokers (10 males and 10 females). Consent was gathered from all participants and the study protocol was approved by the University of Alabama IRB. Subjects were on average 23.1 ± 3.3 years old with an average Body Mass Index of $25.88 \pm 5.24\text{kg/m}^2$ [8].

Each subject performed the following 12 actions: 1) sit silently (5 minutes), 2) read loudly (5 minutes), 3) stand still (5 minutes), 4) walk on a treadmill at a self-selected slow pace (5 minutes), 5) walk on a treadmill at a self-selected fast pace (5 minutes), 6) browse internet on a laptop (5 minutes), 7) eat food without silverware and drink directly from a cup (unrestricted time), 8) eat food with silverware

978-1-4673-7300-5/15/$31.00 ©2015 IEEE

and drink using a straw (unrestricted time), 9) walk outside the building (5 minutes), 10) smoke a cigarette in a sitting position (unrestricted time), 11) rest in a sitting position (5 minutes), and 12) smoke a cigarette in a standing position (unrestricted time).

Experiments were conducted using the machine learning modules within MATLAB's Stats package. Group models were constructed by aggregating data from 19 subjects with the data from the remaining subject held out to validate model performance. That is, 20-1 leave-one-out cross-validation was conducted to estimate the predictive accuracy of group models. Individual models were evaluated using a 5-fold cross validation strategy. That is, data points were divided randomly into 5 subsets. 5 models were constructed by holding out, in turn, each subset or fold and training on the rest of the data observations. Models were then evaluated on the held out fold.

C. Signal Pre-processing

Initially, the proximity sensor ($PS(t)$) signal was normalized to $[0, 1]$. A tidal volume signal was calculated, defined as the average of the $TC(t)$ and $AB(t)$ signals. The amplitude of the $VT(t)$ signal was then scaled to $[-1.0, 1.0]$. In order to reduce signal artifacts, an ideal band pass filter was utilized with cut off frequencies of 0.0001 Hz and 10 Hz. This also eliminated baseline drift. Finally, the $VT(t)$ signal was denoised using a moving average.

The airflow signal is defined as the first derivative of the processed tidal volume $VT(t)$ signal [9] and calculated: $AS(t) = \frac{dV(t)}{dt}$.

D. Feature Extraction

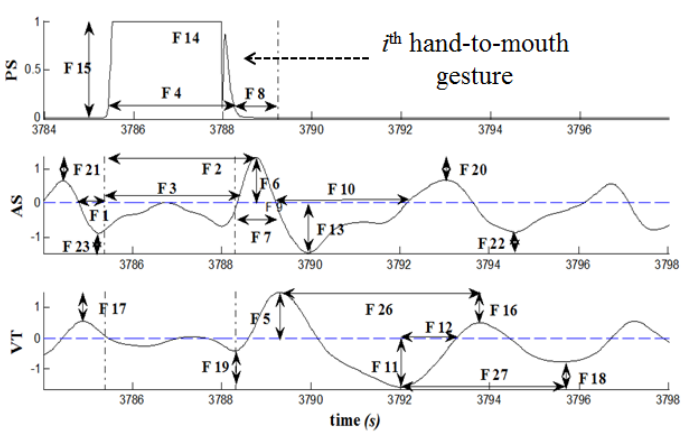

Figure 1. Graphical description of extracted features [10]

27 descriptive features were extracted from each hand-to-mouth gesture and are listed in Table I. These are described in more detail in [10], [8].

E. Models

1) Decision Trees: A decision tree is a predictive model that recursively partitions observations into leaves that are successively more pure (as measured using a variety of methods) [11]. Branches of the tree represent splits of the data by a single variable. The recursive partitioning procedure stops when either the leaf is "pure" (all observations in the node belong to the same class) or another stopping criteria (such as a minimum number of observations per leaf) is satisfied. Gini's diversity index (gdi) was chosen to measure purity of potential splits ($gdi = 1 - \sum_i p(i)^2$).

For example, when predicting customer churn, a variable "called customer service in the last 3 months" that takes value 1 if the customer used the company's customer service system within the previous 3 months and value 0 otherwise may be a useful split. That is, customers who called the customer service are likely to have had product issues and so we may predict that subset to be more likely to switch companies.

Table I
EXTRACTED FEATURES [10]

Feature	Description
1	duration of expiration
2	time-duration from the start time of a hand gesture to peak of $AS(t)$
3	time-duration of hand-to-mouth gesture to point of air-flow exceeding threshold
4	time-duration of hand remaining near mouth
5	peak inspiration following the hand-to-mouth gesture ($AS(t)$)
6	peak inspiration volume following the hand-to-mouth gesture ($VT(t)$)
7	inspiration duration
8	time-duration between the hand leaving the mouth and the beginning of respiration,
9	duration of smoking holding
10	expiration duration ($AS(t)$)
11	breath volume
12	expiration duration ($VT(t)$)
13	maximum expiration level
14	mean amplitude of PS
15	maximum amplitude of PS
16	relative difference between the peak inspiration following a hand-to-mouth gesture and the next peak
17	relative difference between the peak inspiration following the hand-to-mouth gesture and previous peak
18	relative difference between maximum expiration level following the hand-to-mouth gesture and the next breathing cycle's maximum expiration level
19	relative difference between maximum expiration level following the hand-to-mouth gesture and previous breathing cycle's maximum expiration level
20	relative difference between peak inspiration following hand-to-mouth gesture and subsequent peak
21	relative difference between the peak inspiration following the hand-to-mouth gesture and previous peak
22	relative difference between maximum expiration level following the hand-to-mouth gesture and next breathing cycle's maximum expiration level
23	relative difference between maximum expiration level following the hand-to-mouth gesture and previous breathing cycle's maximum expiration level
24	L_2-norm of the smoking breathing cycle and the next breathing cycle ($AS(t)$) (not shown)
25	L_2-norm of the smoking breathing cycle and the next breathing cycle (not shown)
26	time-duration between peak inspiration following hand-to-mouth gesture and next peak
27	time-duration between maximum expiration level following hand-to-mouth gesture and next maximum expiration level.

2) AdaBoost: A boosted classifier is constructed as a weighted sum of base classifiers, denoted $h(x)$, $F(x) = \sum_{i=0}^{T} \alpha_i h_i(x)$ where classification is performed by weighted majority vote. The general AdaBoost (ADAptive BOOSTing) is: at time t a learner is trained on all the data with weights determined by the previous $t-1$ rounds of boosting. That learner is added to the ensemble and the observation weights are updated to minimize the current error. A more thorough description of the algorithm is given in [12]. Models were trained for 50 rounds of boosting.

3) Bootstrap aggregating: From a training set D of size $|D| = n$, bootstrap aggregating or bagging creates m training sets of size n' by sampling from D randomly and with replacement. A model is trained on each of the bootstrapped training sets [13]. Final classification is performed by majority vote of the individual models. In our testing we constructed and trained models on 200 bootstrapped samples of size $n' = .8 \cdot n$.

4) Random Forest: A random forest classifier is a bagged decision tree ensemble with the following modification: at each split in the training procedure, the algorithm selects a random subset (of size $\sqrt{n}$) of the n features. This is done to decorrelate the trees (features that may be dominant due to random chance are less likely to skew all the trees) and thus reduce bias [14]. For each forest, 200 trees were constructed.

F. Model Evaluation

For each fold in the cross-validation scheme, the following performance statistics were computed:

- Precision: $P = TP/(TP + FP)$
- Recall:$R = TP/(TP + FN)$
- F-score: $F = 2 \cdot P \cdot R/(P + R)$

TP is the number of true positives (gestures classified as part of a smoke inhalation that are actually part of a smoke inhalation), FP is the number of false positives (gestures labeled smoke inhalations that are actually another type of gesture), and FN is the number of false negatives (gestures not label to be smoke inhalations that actually are smoke inhalations). In short, precision measures the fraction of positively labeled observations that are truly positive, recall measures the fraction of truly positive observations that were labeled positive, and the F-score is the harmonic mean of precision and recall.

This process was repeated for each subject. Statistics for all models were averaged (and standard deviation computed) to estimate classifier performance on new data.

III. Results

We summarize the results of this study in the following tables:

Table II
Individual ensemble models

Model	F-score (%)	Precision (%)	Recall (%)
Adaboost	77.60 ± 20.40	84.06 ± 10.31	79.08 ± 20.99
Bagging	82.84 ± 10.55	86.82 ± 10.45	83.44 ± 10.60
Random Forest	84.41 ± 10.16	90.18 ± 10.86	84.37 ± 10.06

Table III
Individual SVM models

Model	F-score (%)	Precision (%)	Recall (%)
27 Features (SVM)	68.67 ± 27.28	73.48 ± 24.28	68.38 ± 28.85

Table IV
Group ensemble models

Model	F-score (%)	Precision (%)	Recall (%)
Adaboost	71.66 ± 12.98	80.21 ± 14.01	69.09 ± 19.76
Bagging	70.75 ± 14.23	80.32 ± 13.55	67.01 ± 20.84
Random Forest	69.12 ± 17.63	82.92 ± 13.25	63.06 ± 22.99

Table V
Group SVM models

Model	F-score (%)	Precision (%)	Recall (%)
27 Features (SVM)	65.09 ± 21.64	76.56 ± 17.96	61.32 ± 26.51

The performance of these models is visualized in Figure 2 by displaying the false positive/false negative tradeoffs through several receiver operating characteristic (ROC) curves for Random Forest group models.

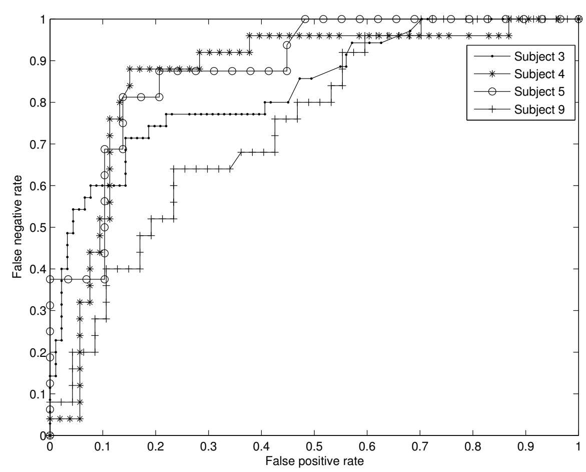

Figure 2. Selected ROC curves for group models with Random Forest classifiers

The training of these models is visualized in Figure 3 by displaying the out-of-bag error for 4 subject's group models as a function of trees in the ensemble.

IV. Discussion

Models of all three types of decision tree ensembles outperformed the SVM classifier using the 27 features (Table II and Table III) on average. Not only did decision tree ensembles outperform in terms of average F-score, precision, and recall, these models also had a lower variance than SVM models on average. Lover variance on average is advantageous in that it allows us to more accurately quantify and describe the performance of these models. Larger variance of results implies that model type (in this situation, SVM) but not

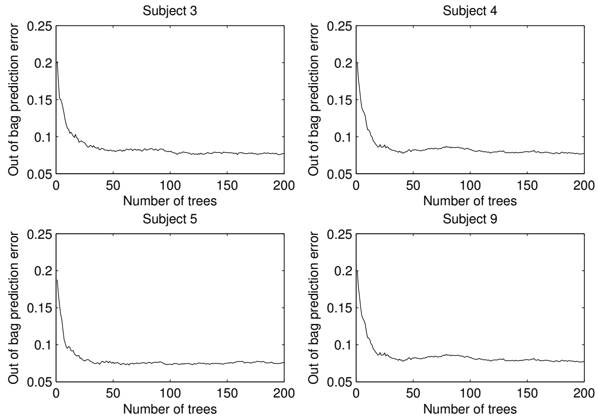

Figure 3. Number of trees vs out-of-bag prediction error for selected subject group models using Random Forest classifier

be as well suited to this problem. Of the individual models trained, Random Forest classifiers performed best across all performance metrics. It is apparent that there is significant individual variation that a group model may not be able to fully capture. This is evident from the higher accuracy that individual models possess (on average).

As we can see from Table IV, AdaBoost outperformed the other two techniques as measured by F-score, but not by a large amount. AdaBoost models also had lower variance on F-scores and recall rates than other model types. Random forests exhibited higher precision rates on average than both AdaBoost and bagging but at the expense of lower recall ability. Average performance (across all model types) was higher than our previous group results computed using the extracted waveforms but lower than results found using 1503-element feature vectors (Table V, [10]).

Figure 2 illustrates the performance of the Random Forest group models by plotting several ROC curves. It is evident that models generalize better to certain subjects than others. This is to be expected and indicates that certain validation subjects are easier, that is, that their breathing patterns are similar to training subjects and that certain subjects' breathing patterns are dissimilar to the rest of the group. In particular, the models generalize better to Subject 3 than Subject 9.

Figure 3 visualizes the performance of models (as measured by out-of-bag error) as a function of the number of trees in the ensemble for the subjects included in Figure 2. It is apparent from this figure that 200 trees probably more than strictly necessary, as generalization error fell rapidly from 1 to 50 trees and then stabilized from 50 to 200 trees.

Some suggestions for future studies include:

1) Obtaining further samples from each individual participant, for example by having the participant repeat the activities on a later trial.
2) Explore the utility of feature selection methods (utilized in previous work [10]) for decision tree ensemble models.

V. Conclusion

Prior to tuning or optimization, decision tree ensemble models outperform previously tested SVM classifiers in individual models on average. The group ensemble models perform comparably, that is, they outperform but only slightly, previously tested SVM models. Models tested provide a method to accurately classify new, never before seen hand-to-mouth gestures. This is a key component of a future system to fully quantify cigarette smoke intake in free living invironments.

References

[1] P. W. Yoon, B. Bastian, R. N. Anderson, J. L. Collins, and H. W. Jaffe, "Potentially preventable deaths from the five leading causes of death–United States, 2008-2010." *MMWR. Morbidity and mortality weekly report*, vol. 63, pp. 369–74, 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24785982

[2] E. J. Perez-Stable, B. Vanoss Marin, G. Marin, D. J. Brody, and N. L. Benowitz, "Apparent underreporting of cigarette consumption among Mexican American smokers," *American Journal of Public Health*, vol. 80, pp. 1057–1061, 1990.

[3] A. A. Strasser, W. B. Pickworth, F. Patterson, and C. Lerman, "Smoking topography predicts abstinence following treatment with nicotine replacement therapy," *Cancer Epidemiology Biomarkers and Prevention*, vol. 13, pp. 1800–1804, 2004.

[4] P. Lopez-Meyer, S. Tiffany, and E. Sazonov, "Identification of cigarette smoke inhalations from wearable sensor data using a Support Vector Machine classifier," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 4050–4053.

[5] E. Sazonov, K. Metcalfe, P. Lopez-Meyer, and S. Tiffany, "RF hand gesture sensor for monitoring of cigarette smoking," in *Proceedings of the International Conference on Sensing Technology, ICST*, 2011, pp. 426–430.

[6] E. Sazonov, P. Lopez-Meyer, and S. Tiffany, "A Wearable Sensor System for Monitoring Cigarette Smoking," *Journal of studies on alcohol and drugs*, vol. 74, no. 6, p. 956, 2013.

[7] P. Lopez-Meyer, Y. Patil, S. Tiffany, and E. Sazonov, "Detection of hand-to-mouth gestures using a RF operated proximity sensor for monitoring cigarette smoking," *Open Biomedical Engineering Journal*, vol. 7, pp. 41–49, 2013. [Online]. Available: http://benthamscience.com/open/openaccess.php?tobej/articles/V007/41TOBEJ.htm

[8] P. Lopez-Meyer, S. Tiffany, Y. Patil, and E. Sazonov, "Monitoring of cigarette smoking using wearable sensors and support vector machines," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 1867–1872, 2013.

[9] A. Eberhard, P. Calabrese, P. Baconnier, and G. Benchetrit, "Comparison between the respiratory inductance plethysmography signal derivative and the airflow signal," in *Frontiers in Modeling and Control of Breathing*. Springer, 2001, pp. 489–494.

[10] Y. Patil, P. Lopez-Meyer, S. Tiffany, and E. Sazonov, "Detection of cigarette smoke inhalations from respiratory signals using reduced feature set," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2013, pp. 6031–6034.

[11] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, 1984, vol. 19.

[12] Y. Freund and R. Schapire, "A short introduction to boosting," *Journal-Japanese Society For Artificial Intelligence*, vol. 14, no. 771-780, p. 1612, 1999.

[13] L. Breiman, "Bagging predictors," pp. 123–140, 1996.

[14] ——, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.