Final Report: Classification of Behaviors During Eating Using Data Obtained from Multiple Sensors

(Chris) Yaru Niu
AICPS Lab, University of California, Irvine
South China University of Technology
yarun@uci.edu, auchrisniu@mail.scut.edu.cn

***Abstract*— Monitoring food intake is an important approach to help people keep a balanced diet thus maintain a healthy liftstyle. Nowadays, health internet of things (IoT) systems, usually employing wearable equipments with sensors, are developed and applied to monitoring eating behaviors, analyzing eating habits and generating recommendations of people's diet. To monitor the behaviors during eating, time-sequence data, from which we label the time step with eating behaviors, are extracted from multiple sensors including piezoelectric sensor, strain sensor and microphone. First, considering the continuity and correlation of different time steps in the time-sequence data, the features of the data are extracted using sliding windows, then the feature vectors and the corresponding labels are used to train the classifiers including Random Forest, K-Nearest Neighbors, Decesion Trees, Naive Bayes, Adaboost, Support Vector Machine and Deep Neural Networks. Therefore, given the feature vector, we can obtain the predicted label using the trained model at each time step. Thereafter, the mode filter sliding window is employed to alleviate the prediction noises. In this work, the trained models are verified by the K-fold cross-validation, which indicates Random Forest has the best prediction accuracy.**

***Key Words*— Eating Behavior Monitoring, Feature Extraction, Supervised Learning, Classification, Sliding Window**

I. Introduction

When medical treatments and pharmaceutical drugs are invested and imporved to cure various diseases, preventative healthcare should be emphasized to prevent diseases and save money [10]. Besides public health measures and encouragement of exercise, quantifying diet is another effective approach for preventative healthcare [7]. Various technologies have been applied to dietary monitoring such as maunual record [15], audio-based analysis [5], [8], [13], [14], gesture recognition [16], camera-based techniques [11], and piezo or strain detection [4], [6]. Based on the dietary monitoring, the eating habits can be expected to analyze and recommendations of people's diet can be expected to be given.

Recognizing and classifying behaviors during eating is an important task of dietary monitoring. Kaltantarian et al. recognized the swallowing by detecting the peaks of the processed piezo data and classifying other behaviors including looking up, walking and turning during eating using the accelerometer data [4], [6]. Audio-based methods often extract features from the frequency spectrum of the audio data [5], [14], and the behaviors during eating can be classified by some machine learning algorithms such as Support Vector Machine, Naive Bayes and Random Forest.

In this report, we present a method of classifying the behaviors during eating through the data obtained from the piezoelectric sensor, strain sensor and microphone. Sliding window technologies have been used in some health monitoring systems [4], [6], [14], and other work can be found in [2], [3]. In our work, the means and standard deviations of each time slip are extracted as the features by sliding windows, then the feature vectors and the corresponding labels are used to train the classifiers including Random Forest, K-Nearest Neighbors, Decesion Trees, Naive Bayes, Adaboost, Support Vector Machine and Deep Neural Networks. After given the feature vector, the trained model can be used to generate the predicted label at each time step. Then the mode filter sliding window is employed to eliminate the prediction noises. At last, the trained models are verified by the K-fold cross-validation respectively.

II. Methodology

In our work, the data are collected from the piezoelectric sensor, strain sensor and microphone. The sliding windows are applied to extracting features, then several machine learning algorithms are employed to predict the labels. Thereafter the mode filter is used to alleviate the noises of the predicted labels. At last our method is verified by the K-fold cross-validation. The methodology will be discussed in detail in this section.

A. Data Collection

To collect the raw data, the volunteers are instructed to act a series of behaviors including opening mouth, chewing nuts, swallowing nuts and talking. Meanwhile, the piezoelectric sensor, the strain sensor and the microphone are attached to the subject, and their data stream are recorded in a laptop. The data are labeled by pressing the button on the user graphical interface.

B. Feature Extraction

To extract the features for classification, we use the sliding window on the time-sequence data of piezo, strain and microphone, respectively, and calculate the mean and the standard deviation value in each window. Because we have various behaviors to classify, and the durations of each behavior slip are different, some of which are close to the window size, we apply the sliding window with a maximum overlap which means the window shift one time step at a time. The calculated mean and standard value are put into the centering position of each window.

C. Classification Models

To predict the data's labels, a specfic classification model needs to be constructed and trained to which the feature vectors are input. We choose Random Forest, K-Nearest Neighbors, Decesion Trees, Naive Bayes, Adaboost, Support Vector Machine and Deep Neural Networks as the classification models and compare their performances. For the former 6 models, we use Scikit-learn, which is a Python module for medium-scale supervised and unsupervised problem [12]. To construct, train a neural network with 5 classes and 2 hidden layers each of which has 10 neurons, Tensorflow [1] is employed. Compared to Scikit-learn, Tensorflow is a large-scale system for machine learning problems.

D. Mode Filter

After getting the predicted results from the machine learning algorithms, the mode filter is used for eliminating the noises and improving the performance of the estimator. The filter is based on a sliding window of length 51. In each window, the mode of the label numbers are calculated and set as the value of the center of the window.

E. Cross-Validation

Cross-validation is a commonly-used technique to evaluate the performance of the machine learning algorithms, which holds out part of the available data as the test set iteratively. K-fold cross validation divides all the samples in k mutually exclusive groups of approxiamately equal size [9]. The prediction model is learned by training the $k-1$ groups, and the group left out is used for testing.

F. Evaluation Metrics

In our work, the classifiers are evaluated by accuracy, precision, recall and F-measure respectively, which are defined as follows:

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+FP+FN+TN}, \\ Precision &= \frac{TP}{TP+FP}, \\ Recall &= \frac{TP}{TP+FN}, \\ F-Measure &= \frac{2*Precision*Recall}{Precision+Recall} \\ &= \frac{2TP}{2TP+FP+FN}, \end{aligned} \tag{1}$$

where TP, TN, FP, FN represent true positives, true negatives, false positives and false negatives, respectively. As we have 5 classes (Opening Mouth, Chewing, Swallowing, Talking and Other), the metrics are calculated for each class, and the final score is obtained by calculating their average weighted by the number of true instances of each class.

III. Results

We empolyed the dataset with the size of 26286 collected from 5 subjects to do 50-fold (50 groups) cross-validation using the 7 machine learning algorithms. The size of the sliding window for feature extraction was set as 33, and the window size of the mode filter was set as 51. Because the value of the window was put into the center, and the sizes of the sliding windows for feature extraction and mode filter were 33 and 51 respectively, 41 samples of data were eliminated at both the start and the end of the time-sequence dataset. The results of evaluation metrics for each algorithm are shown in Table 1.

TABLE I
PERFORMANCE EVALUATION OF EACH ALGORITHM

Algorithm	Accuracy	Precision	Recall	F-Measure
Random Forest	77.05%	77.72%	77.05%	77.25%
Deep Neural Networks	70.16%	70.58%	70.16%	68.00%
Decision Trees	70.00%	71.68%	70.00%	70.23%
K-Nearest Neighbors	68.82%	69.04%	68.82%	68.82%
AdaBoost	59.09%	65.12%	59.09%	60.32%
Naive Bayes	55.27%	62.93%	55.27%	56.02%
Support Vector Machine	23.16%	29.11%	23.16%	9.10%

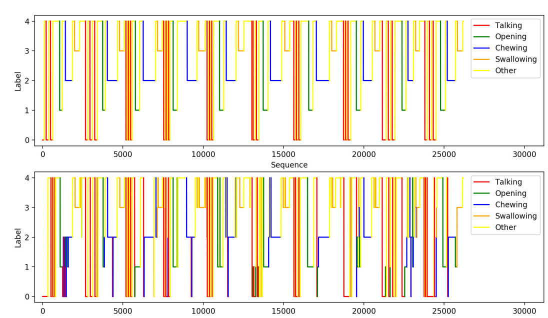
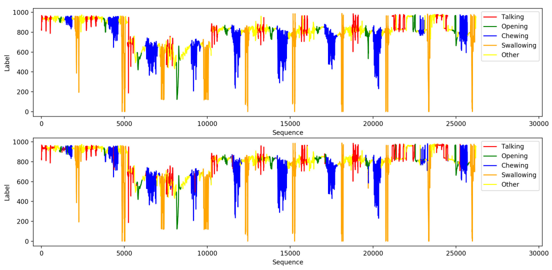
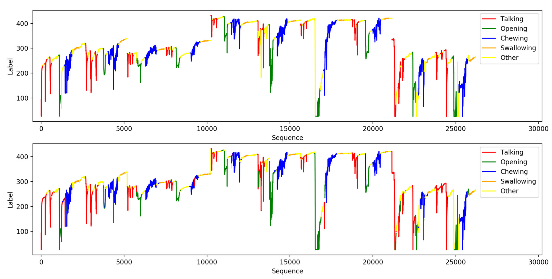


Fig. 1. The actual labels and the predicted labels generated by Random Forest through the cross-validation iterator. "Talking" is colored by red, "Opening" is colored by green, "Chewing" is colored by blue, "Swallowing" is colored by orange, and "Other" is colored by yellow.

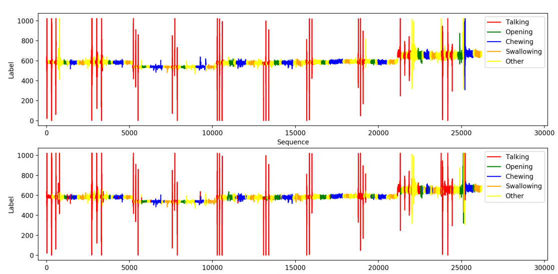
As can be seen from Table 1, Random Forest has the best performance of our classification problem. The actual labels and the predicted labels generated by Random Forest through the cross-validation iterator are depicted in Fig. 1. The piezo, strain and microphone data colored by colors which represent the actual and predicted labels are depicted in Fig.2 (a), (b) and (c), respectively. Other than Random Forest, Deep Neural Networks and Decision Trees also have relatively good performance. On the hand, the performance of Support Vector Machine is bad. This might be due to the high-dimensional feature space and output space.



(a) The colored piezo data.



(b) The colored strain data.



(c) The colored microphone data.

Fig. 2. The piezo, strain and microphone data are colored by the colors which represent the actual and the predicted labels of Random Forest. "Talking" is colored by red, "Opening" is colored by green, "Chewing" is colored by blue, "Swallowing" is colored by orange, and "Other" is colored by yellow.

IV. Conclusion

In this work, we use the piezoelectric sensor, strain sensor and microphone to collect the time-sequence data of people's behaviors during eating. The maximum-overlapped sliding window is employed to extract the features of the data. Then the features with labels can be used to train the classification models, which is applied to predicting the labels of the newly unlabeled data. To eliminate the prediction noises, we use the mode filter to process the time-sequence labels. At last, the 7 classification models are compared and verified by the K-fold cross-validation, which indicates Random Forest has the best performance.

References

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.

[2] Yun Chi, Haixun Wang, Philip S Yu, and Richard R Muntz. Moment: Maintaining closed frequent itemsets over a stream sliding window. In *Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on*, pages 59–66. IEEE, 2004.

[3] Lilian Harada. An efficient sliding window algorithm for detection of sequential patterns. In *Database Systems for Advanced Applications, 2003.(DASFAA 2003). Proceedings. Eighth International Conference on*, pages 73–80. IEEE, 2003.

[4] Haik Kalantarian, Nabil Alshurafa, Tuan Le, and Majid Sarrafzadeh. Monitoring eating habits using a piezoelectric sensor-based necklace. *Computers in biology and medicine*, 58:46–55, 2015.

[5] Haik Kalantarian, Nabil Alshurafa, Mohammad Pourhomayoun, Shruti Sarin, Tuan Le, and Majid Sarrafzadeh. Spectrogram-based audio classification of nutrition intake. *IEEE EMBS Healthcare Innovations & Point of Care Technologies (HIPT)*, 2014.

[6] Haik Kalantarian, Nabil Alshurafa, and Majid Sarrafzadeh. A wearable nutrition monitoring system. In *Wearable and Implantable Body Sensor Networks (BSN), 2014 11th International Conference on*, pages 75–80. IEEE, 2014.

[7] Haik Kalantarian, Nabil Alshurafa, and Majid Sarrafzadeh. A survey of diet monitoring technology. *IEEE Pervasive Computing*, (1):57–65, 2017.

[8] Haik Kalantarian and Majid Sarrafzadeh. Audio-based detection and evaluation of eating behavior using the smartwatch platform. *Computers in biology and medicine*, 65:1–9, 2015.

[9] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.

[10] Michael V Maciosek, Ashley B Coffield, Thomas J Flottemesch, Nichol M Edwards, and Leif I Solberg. Greater use of preventive services in us health care could save lives at little or no cost. *Health Affairs*, 29(9):1656–1660, 2010.

[11] Jon Noronha, Eric Hysen, Haoqi Zhang, and Krzysztof Z Gajos. Platemate: crowdsourcing nutritional analysis from food photographs. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 1–12. ACM, 2011.

[12] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[13] Andrea Santamato, Francesco Panza, Vincenzo Solfrizzi, Anna Russo, Vincenza Frisardi, Marisa Megna, Maurizio Ranieri, and Pietro Fiore. Acoustic analysis of swallowing sounds: a new technique for assessing dysphagia. *Journal of rehabilitation medicine*, 41(8):639–645, 2009.

[14] Edward S Sazonov, Oleksandr Makeyev, Stephanie Schuckers, Paulo Lopez-Meyer, Edward L Melanson, and Michael R Neuman. Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior. *IEEE Transactions on Biomedical Engineering*, 57(3):626–633, 2010.

[15] Dale A Schoeller. How accurate is self-reported dietary energy intake? *Nutrition reviews*, 48(10):373–379, 1990.

[16] Sougata Sen, Vigneshwaran Subbaraju, Archan Misra, Rajesh Krishna Balan, and Youngki Lee. The case for smartwatch-based diet monitoring. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, pages 585–590. IEEE, 2015.