Information Fusion 41 (2018) 37–47

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

I sense overeating: Motif-based machine learning framework to detect overeating using wrist-worn sensing

Shibo Zhang[a,*], William Stogin[a], Nabil Alshurafa[b]

[a] EECS, Northwestern University, USA
[b] Preventive Medicine Dept. & EECS, Northwestern University, USA

ARTICLE INFO

Article history:
Received 4 February 2017
Revised 19 June 2017
Accepted 3 August 2017
Available online 4 August 2017

Keywords:
Wrist-worn sensors
Wearables
Hand-to-mouth gestures
Overeating
Inertial sensors
Motif-based segmentation
K-Spectral Centroid Clustering
Fusion
Classification
Feeding gesture

ABSTRACT

Obesity, caused primarily by overeating, is a preventable chronic disease yielding staggering healthcare costs. To detect overeating passively, a machine learning framework was designed to detect and accurately count the number of feeding gestures during an eating episode to characterize each eating episode with a feeding gesture count using a 6-axis inertial wrist-worn sensor. Moreover, detecting feeding gestures is useful to aid in end-of-day dietary recalls. It has been shown that feeding gesture count correlates with caloric intake; the more one eats, the more calories one is likely consuming. Recent research has shown promise in passively detecting feeding gestures, but this effort focuses on bridging detection of feeding gesture count and identifying overeating episodes. This paper presents results on three experiments: highly structured (participants pretending to eat), in-lab structured with confounding activities (participants eating while performing other scripted activities), and unstructured overeating (participants induced to overeat while watching television and eating their favorite foods). Our experiment successfully induced overeating in 50% of the participants, showing a correlation between feeding gesture count and caloric intake in unstructured eating (r=.79, p-value=.007). Results provide an approximate upper bound on feeding gesture classification using exact segmentation techniques, and show improvement when compared to prior sliding window techniques. Results also suggest the importance of stressing the challenge of accurate segmentation over identifying the accurate classification technique in detection of feeding gestures. Since participant-dependent models provide optimal results, a motif-based time-point fusion classification (MTFC) framework is proposed using spectral energy density, K-Spectral Centroid Clustering, symbolic aggregate approximation (SAX), a Random Forest classifier (trained on segmented motifs) and a time-point classifier fusion technique to show reliable classification of feeding gestures (75% F-measure), and a 94% accuracy of feeding gesture count in the unstructured eating experiment, resulting in a root mean square error of 2.9 feeding gestures. Mapping feeding gesture count to caloric intake, we obtain a rough estimate of whether participants overate while watching television.

1. Introduction

Eating is essential to human life, but overeating relative to need is not. Unfortunately, once bad eating habits are formed, they become challenging to overcome. People overeat for many reasons, such as loss of control [1], impulsivity due to cues [2], or heightened emotional state as a result of stress [3] or negative affect [4], or even positive affect [5]. Being able to passively detect overeating in real time will enable researchers to understand the antecedents and causes of overeating.

Existing studies have already demonstrated the feasibility of detecting eating behavior [6]. One phenomenon in feeding gesture detection is the large variability in performance of detection feeding gestures among works. One possible explanation of the variance is that performance is not only dependent on algorithm design but also significantly dependent on the quality of the dataset, including but not limited to factors such as if the participant is conducting a standard gesture, how many feeding gestures are included in an eating episode as well as how many non-feeding gestures are included.

Our objective is to use wearable sensors not only to detect eating episodes and aid in dietary recall, but also to identify problematic eating episodes, where caloric overeating is likely to occur. We hypothesize that eating duration as well as the number of feeding gestures, swallows and chews are important physical

* Corresponding author.
E-mail addresses: shibozhang2015@u.northwestern.edu, zhangshibo87@gmail.com (S. Zhang).

http://dx.doi.org/10.1016/j.inffus.2017.08.003
1566-2535/

Table 1
List of terms and descriptions.

Term	Description
FG	Feeding gestures
nFG	Non-feeding gestures
# FG	The number of feeding gestures from ground truth
# FG prediction	The number of feeding gestures from prediction
# FG accuracy	The accuracy of number of feeding gestures from prediction

features that can characterize overeating. This effort focuses on a framework for characterizing challenging eating episodes through feeding gestures using wrist-worn inertial sensors. Challenges in this effort stem from two major aspects: (1) a single person can have feeding gestures with a variety of feature representations resulting from different utensils, miscellaneous table manners, improvisational feeding behaviors and unexpected sudden events like food and utensil dropping; (2) different personal behavior habits (stroking one's hair, scratching one's face) resulting in different eating rates, which makes it challenging to fit a generalized model to all users.

Due to the challenges described above, and since performance of feeding gesture classification depends on the test protocol and the behavior of each participant we explore the classification ability of machine learning algorithms to distinguish feeding gestures from feeding-like activities. We designed and conducted three different experiment protocols, including a highly structured test (participants pretending to eat), in-lab structured test with confounding activities (participants following an eating protocol), and an unstructured test (participants induced to overeat while watching T.V. and eating their favorite foods, after being full).

To detect and count feeding gestures, we designed a framework in two steps. First, we extracted motifs from exact segments (defined by ground truth) to build a database of motifs. Secondly, we perform K-Spectral Centroid Clustering to extract motif templates and perform motif matching to search for candidate feeding gesture segments. Thirdly, motif matching is performed using a symbolic aggregate approximation (SAX) method, followed by feature extraction, Random Forest classification, and a unique decision-level fusion method combining majority voting [7] and multiple overlapping window segments to detect and count feeding gestures.

The main contributions of this paper comprise:

1. a new time-point classifier fusion framework using motif-based templates with varying window lengths to count feeding gestures;
2. an upper bound of feeding/non-feeding gesture classification using exact segmentation;
3. a comparison between a motif-based segmentation technique and a previous sliding window technique;
4. promising results for predicting overeating behavior with only one wrist-worn motion sensor. Through detecting and counting the feeding gestures we can effectively predict overeating intervals; and
5. a well-labeled and video-recorded dataset containing a range of three experiment protocols.

Table 1 summarizes notations used in reporting results in this paper.

2. Related works

Sensing devices that use three-axis accelerometers and three-axis gyroscopes to define eating gestures and activities come in many forms and in many body locations including the wrist, neck, arm, and trunk [8–16]. Table 2 summarizes existing literature on detecting eating using feeding gesture counts along with the sensors used, segmentation methods and test data set specifications. Our results improve on existing literature by testing on more confounding gestures in lab and induction of T.V. watching-linked overeating episodes, while not compromising accuracy.

The widespread availability of embedded wearable accelerometers and gyroscopes has enabled a new area of research to detect eating passively through on-body inertial sensors. The focus of this research effort is on detecting and characterizing eating through feeding gestures.

The problem of identifying hand-to-mouth gestures has also been studied to detect smoking activities in order to predict smoking relapse [17–19]. PuffMarker successfully detected the timing of a relapse using multiple sensors, where detecting hand-to-mouth gestures and hand orientation (using roll and pitch angles) are part of the smoking lapse system [17]. RisQ also leveraged multiple inertial measurement units (IMUs) placed on a person's body together with 3D animation to detect hand-to-mouth gestures, and while the focus of these systems is primarily smoking, their systems have shown preliminary success in detecting feeding gestures.

Dong et al. showed correlation between bites and caloric intake and measures intake via automated tracking of wrist motion [20,21]. Some of their limitations include requiring the user to turn the device on and off, and they focus primarily on detecting the start and end of an eating episode throughout the day, as opposed to characterizing a given eating episode.

Thomaz et al. presented a framework for detecting eating episodes using wrist-worn accelerometer data [9] and density-based spatial clustering of applications with noise to identify eating episodes throughout the day. The effort focused more on detecting eating episodes that are at-least five minutes long, but our effort focuses on characterizing these eating episodes to obtain a more fine-grained understanding of eating behavior. We also compare our method with the sliding window technique to show improvements in feeding gesture classification.

Amft et al. [8,14] integrated commercial motion sensors in a jacket to capture the movement of the participants wrists, lower arms, upper arms and upper back. Then they built a feeding gesture detection system with the collected data. In comparison, the system that we used for our study is more accessible using a single wrist-worn sensor mounted on the participant's dominant hand.

Although many studies focus on eating gesture detection with various methods and systems, most of those studies are based on fixed-length sliding window segmentation and eating/non-eating gesture classification. Fixed-length sliding window segmentation benefits from simplicity, speed, and robustness. For example, in [9], the author presented presents a fixed-length sliding window machine learning framework to detect and count feeding gestures. In [10,12,13,15], the fixed-length sliding window segmentation framework was adopted. However, there are other segmentation methods. In [22], the author developed an interesting method to extract candidate windows. A fast (0.8 s window) and a slow (8 s window) sliding window were used for calculating the average gyroscope magnitude. Then segments where the fast-moving average lies below the slow-moving average were selected. In these regions, the magnitude is lower than the average magnitude of the neighborhood.

To distinguish our project from existing works, we present a flexible length sliding window approach. Our approach is based on capturing the complete hand up and hand down gesture, resulting in robust and intelligent segmentation and classification of feeding gestures.

Motif matching is the process of finding a specific motif within a signal. It is a wide topic of time series data mining which has been intensively studied in the academic community of bio-

Table 2
Literature on feeding detection system.

Reference	Goal	Sensors	segmentation method	# FG types	# nFG types	# FG accuracy	disadvantage
Junker et al. [8]	Detect eating gestures	(Accel & gyro)*5	flexible-length (SWAB)	–	–	–	–
Thomaz et al. [9]	Detect eating episodes	Accel*1	fixed-length sliding win	–	–	–	–
Dong et al. [10]	Detect eating episodes	(Accel & gyro)*1	fixed-length sliding win	–	–	–	–
Amft et al. [11]	Detect and count feeding gestures	(Accel & gyro)*4	flexible-length (SWAB)	4	1	0.87	Multiple sensors
Maramis et al. [13]	Detect and count feeding gestures	Accel*1	fixed-length sliding win	1	0	0.92	Restriction on gesture duration
Zhang et al. [15]	Detect and count feeding gestures	(Accel & gyro)*1	fixed-length sliding win	5	3	0.87	Parameter-dependent
Sen et al. [16]	Detect and count feeding gestures	(Accel & gyro)*1	fixed-length sliding win	3	5	0.83	Strong heuristics

medical and text data mining. Among numerous proposed algorithms, SAX [23] is probably the most widely used and acknowledged method. By translating time series data into discrete symbolic series, the SAX algorithm significantly decreases the dimension of the signal, thus reducing required computational cost and improving matching speed. It has shown promise in a real-time data streaming application.

3. Devices and experiments

In the following sections we describe our system and methodology in more details.

3.1. Wrist-worn sensor and mobile phone system

An Android application was designed to act as the information gateway between the wristbands and our back-end database. The data is stored in a Couchbase database, and can both be transmitted to a back-end or accessed directly from the smart phone through local storage. The application works with the server to create unique ids for each phone, sensor and participant. From the application, the researcher is able to turn on a combination of sensors (accelerometer, gyroscope, heart-rate, skin temperature, etc.) and adjust their corresponding sample rates. The local database makes it possible to readily upload the data to a back-end database where the data is analyzed for studies in free-living populations.

One essential challenge was streaming data disconnecting from the Microsoft Band 2. The Band2's Application Programming Interface (API) does not alert the phone when the bluetooth connection to the Band is lost. To counteract this, the Android application checks to make sure that data is still streaming every minute, and re-establishes the connection if necessary. If connection cannot be reestablished, the app notifies the user of the lost connection.

3.2. Data collection and video labeling

Three experiments were performed to achieve an upper-bound on classifier accuracy of feeding gestures in structured and unstructured activities. To achieve this upper bound, ground truth is used to extract feeding gestures. Three trained researchers are used to correctly label the video collected into feeding and non-feeding gestures. Participants (10 in total, 1 female and 9 males, mean age=24.2 years) were requested to wear a wrist-worn sensor (Microsoft Band 2) on their dominant hand. A Logitech C615 HD webcam camera was placed on the side of the non-dominant hand in order to see the plate, both hands and the participants face. Chronoviz [24] was used to label data. Although this paper primarily focuses on detecting feeding gestures, the variety of the labels include start and end times for: eating, drinking, feeding gesture (hand to mouth and back to food), and non-feeding gesture (hand up, and back down but not feeding related). All the participants were right-handed, although a few did consume food with their left hand during unstructured eating, and as in this study we do not capture the participants' use of their non-dominant hands, feeding gestures performed with the left-hand represent uncaptured data. The entire session was recorded using a camera. The annotated recordings serve as ground truth for our models. This dataset will be made public to advance feeding gesture detection. Fig. 1 illustrates the system used to annotate the recorded data with time-synchronized ground truth labels.

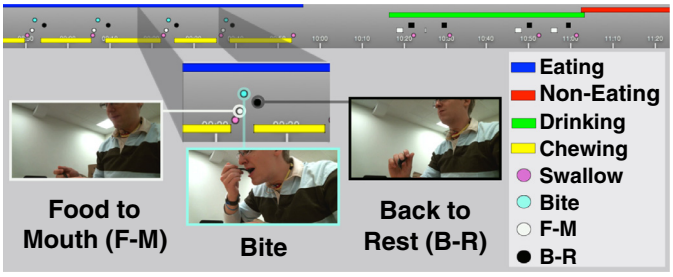

Fig. 1. Labeling data using Chronoviz to enable the detection of feeding gestures.

Table 3
List of activities and descriptions.

Activity	Description
Soup w/ spoon	Eating soup with spoon
Bread w/ hand	Eating bread with hand
Drink w/ cup	Drinking water from a cup
Eat w/ chopstick	Eating salad with chopsticks
Knife & fork	Eating BBQ chicken with knife and fork
Drink w/ bottle	Drinking from a water bottle
Salad w/ fork	Eating salad with fork
Chips w/ hand	Eating chips with hand
Drink w/ straw	Sipping water from a straw
Scratch	Scratching head and face
Chin rest	Resting chin on propped up hands
Phone call	Making a phone call
Smoke w/ thumb and index	Pretending to smoke a cigarette with thumb and index finger
Smoke w/ middle and index	Pretending to smoke a cigarette with middle and index fingers
Comb hair	Combing hands through hair in front of a mirror
Brush teeth	Brushing teeth

3.3. Highly structured test

In the highly structured test, participants are asked to perform 10 gestures of each of the activities described in Table 3 using only a utensil and imaginary food (no food is presented in this experiment). The participants are asked to follow a script instructing them when to perform each gesture, separating each gesture by 10 s of motionless activity. This makes it possible for a simple script to filter out the motion into perfect feeding gesture seg-

Table 4
List of intake calories and feeding gestures for each participant in in-lab structured tests.

Subject	Calories	# FG
1	987.5	101
2	1050	226
3	812	113
4	959	220
5	844	73
6	808	139
7	680	187
8	900	136
9	1030	116
10	1100	120

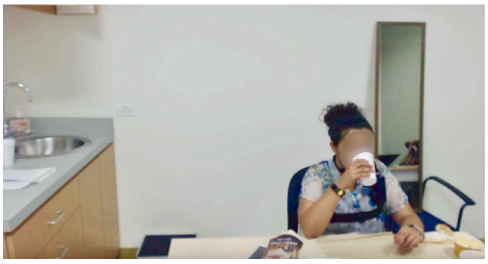

Fig. 2. A participant in the in-lab structured test.

ments. This is essential to see if food is necessary to train a model, or whether pretend eating is sufficient.

3.4. Structured test

In the in-lab structured test, participants performed all the activities (eating and non-eating) described in Table 3 (randomized for each participant) for exactly two minutes each. The participants were instructed to follow a script on a developed interface, which reminded them of the current activity, and instructed them to consume the foods as freely as possible. This test reflected more natural feeding gestures from each participant. Caloric intake was calculated by weighing food before and after the experiment and the number of feeding gestures was obtained from the video. The number of calories and the number of gestures were recorded for each participant (See Table 4). At the end of the experiment participants were asked about the foods they like to eat and what T.V. show or movie they would like to watch. We used that information to tailor the unstructured experiment for each participant. Fig. 2 shows a participant in the in-lab structured test.

3.5. Unstructured test

Participants were asked to return to the lab on another date in the evening to eat food in a lab setting. After putting on the sensors, they were asked if they were full; if they were not full, they were provided with a meal of their choice to eat their fill, then they were taken to a room with a big screen T.V. to watch their favorite T.V. show/movie. If they were full, then they were led directly to the T.V. room. While watching T.V., participants were presented with their favorite snacks (identified in in-lab structured test). During both the pre-load and snacking phases participants provided a large variety of gestures such as unwrapping food and finger licking that challenged the robustness of our algorithms. Caloric intake and the corresponding feeding gesture count for the unstructured test is listed in Table 5. Fig. 3 shows a participant watching T.V. and performing the unstructured test.

Table 5
List of intake calories, 1/2 Harris Benedict, overeat condition, and feeding gestures for each participant in unstructured test.

Subject	Calories	# FG	1/2 Harris Benedict	Overeat
1	180	48	822.5	N
2	1048	117	971.125	Y
3	1642	91	973.25	Y
4	828	104	1078.125	Y
5	895	146	1088.125	N
6	1226	123	970.625	Y
7	1813	163	1050.625	Y
8	1408	224	963.125	Y
9	215	5	1072.5	N
10	95	11	945	N

Fig. 3. A participant in the unstructured test.

3.6. Defining overeating episodes: Harris Benedict equation

The Harris-Benedict (HB) principle is used to estimate basal metabolic rate (BMR) and daily Calorie requirement. However, there currently exists no standard definition of an overeating episode. Assuming participants ate at least two meals a day, participants were considered to exhibit an overeating episode if their caloric intake exceeded 50% of their daily Calorie requirement, i.e. they consumed a large meal in relation to their daily need. The following equations are used to calculate daily Calorie requirements (assuming little to no exercise during that day, and we made sure they were inactive for the day of their unstructured eating test) [25].

$$BMR_{men} = 66 + 6.2W + 12.7H - 6.76A$$
$$BMR_{women} = 655.1 + 4.35W + 4.7H - 4.7A$$
$$Calorie = BMR_x * 1.2$$

where BMR_x represents BMR calculation for x=men, women, W is weight in pounds, H is height in inches, and A is age in years.

We list the calculated 1/2 Harris Benedict for each participant and whether they overate or not along with their feeding gesture count in Table 5.

3.7. Feeding gestures and calories

Generally speaking, ones caloric intake often increases with more food consumed. Although with different food categories the calories in one bite may vary, the general trend of more feeding gestures indicating higher caloric intake remains consistent [20,26]. In prior efforts, we showed that feeding gestures correlate with caloric intake, and we further demonstrate this in the unstructured eating episode.

The Pearson product-moment is used to examine the association between kilocalories and the number of feeding gestures. The Spearman rank-order correlation evaluates the monotonic relation-

Table 6
In-lab unstructured test correlation calculations between calories and count of feeding gestures.

Variable	Spearman	Pearson
Correlation coefficient	0.72	0.79
P-value	0.019	0.007

ship between two continuous or ordinal variables. Table 6 shows that the correlation during unstructured eating episodes also holds true, which indicates that feeding gestures may be an important contributor to determining an estimate of caloric intake.

4. Methodology

Data segmentation is the process that identifies segments of data streams that likely contain information about the activity (e.g. feeding gesture) performed. In an activity recognition tool chain, this can be one of the greatest challenges. Segmentation approaches generally fall under three methods: sliding window, energy-based (e.g. rest position can be a certain posture), or using additional sensors or context (e.g. a proximity sensor detecting the hand close to the mouth). We chose the scheme that utilizes only one smart wristband with a motion sensor mounted on a participants wrist, because of our emphasis on adherence and wearability. In this work we selected an energy-based segmentation method and compare the effects of classification using an energy-based segmentation and a fixed-length sliding window segmentation.

4.1. Upper bound on feeding gesture classification

Using Chronoviz, we perform human annotation of feeding gestures and confounding non- feeding gestures (which show similar characteristics to feeding gestures, including hand-up and hand-down motions). Using ground truth segments, we test multiple classification algorithms ability to distinguish between feeding and non-feeding gestures. This enabled us to obtain an upper bound on our feeding gesture classification (assuming we can detect hand-up and hand-down motion). We tested nine widely adopted machine learning algorithms that build generative and discriminative models from the training set including Nearest Neighbors, Gaussian Process, Decision Tree, Random Forest, Neural Net, AdaBoost, Naive Bayes, QDA, and Linear SVM. We tested a personalized model and generalized models. Personalized models were designed using a 70%/30% train/test split ratio. For generalized models, we tested 10-fold cross validation, Leave One Participant Out (LOPO), and a 70%/30% train/test split ratio. Among these classifiers, Random Forest achieved optimal results, which is consistent with prior literature [9,26]. When splitting the training and test sets we ensured every activity identified in Table 3 was present in both the training and test sets. Results provide insight into the best possible performance given ideal segmentation.

4.2. Data collection

We used a Microsoft Band 2 to collect accelerometer and gyroscope data at a sampling rate of 31 Hz. Microsoft Band 2 provides frequency options of 8 Hz, 31 Hz and 62 Hz, and prior literature [26] found 31 Hz to be sufficient resolution in capturing feeding gesture counts. During some of the experiments, we found that the smart watch automatically started collecting data at 8 Hz instead of 31 Hz. Given initial analysis, we found that data collected at 8 Hz was sufficient for classification given exact segmentation, however it was not sufficient when using our proposed auto-mated motif-based segmentation approach.

4.3. Data preprocessing

We show the overall architecture of our motif-based segmentation and classification in Fig. 4. Motif templates are obtained from the training set and then used for segmentation in the training and test sets. Machine learning models are trained with the ground truth of the training set segments. Then with the trained model, feeding gesture segments are classified and recognized. After we have the segment prediction, we finally apply a time-point fusion method to count feeding gestures.

Initially, we removed any data where the video recorded shows time jumps (missing video data), which occurred when we stored the streaming video data for some participants on an external hard drive. This was a necessary edit since we cannot validate what was performed by the participant without the corresponding video, and removing this data ensures that the inertial signals intended measurements were captured, primarily by smoothing to reduce noise and normalization. The premise of smoothing data is that one is measuring a variable that is both slowly varying and corrupted by random noise. Consequently, replacing each data point with an average of surrounding points reduces the level of noise, while hopefully not biasing the values. To smooth the data we used a recursive low-pass filter with formula $y(n) = 0.3x(n-1) + 0.7y(n-1)$ [27], and to normalize, we used a Z-score normalization relative to sample mean and standard deviation.

4.4. Motif-based segmentation

To generate motifs, we needed to identify a signal or subset of signals from which a motif is generated to represent feeding gestures. We selected the dynamic energy of acceleration as the signal source. The energy of acceleration represents the intensity of ones hand movement. During feeding gestures, with movement of the hand up or down, the intensity of acceleration experienced similar patterns. For feeding gestures with different utensils, the motifs have different shapes. With this motif-based segmentation method, many feeding-like gestures which experience the same intensity change will also be detected.

4.4.1. Defining dynamic energy signal

We define the dynamic energy of the accelerometer signal $E(acc)$ in the following equation. $E(acc)_i$ is the dynamic energy of accelerometer signal at time point i. $X_{x,i}$, $X_{y,i}$, and $X_{z,i}$ are separately the amplitudes of the frequency component of x-, y- and z-axis accelerometer signals. N is the Fast Fourier Transform (FFT) window size. F_s is the sampling rate. The linear acceleration sensor we used is influenced by gravity. Gravity will have an additive effect on the sensor's three-axis output, so as a result, we remove the fundamental frequency component so that n starts from 2.

$$E(acc)_i = \sum_{a=x,y,z} \sum_{n=2}^{N} X_{a,i}^2 \left(f = \frac{(n-1)F_s}{N} \right)$$

4.4.2. Motif generation based on K-Spectral Centroid Clustering

Since we have many kinds of feeding gestures with different utensils, selecting a powerful clustering method to generate distinctive motifs is of essential importance for this study. We applied the time series clustering method K-Spectral Centroid Clustering (KSC) [28] to generate the common motifs for all feeding gestures. KSC method groups a number of motifs and returns the generated centroid of each group, generating several centroids or motif templates. It is an averaging technique that is based on matrix decomposition to compute the centroid of a cluster when a distance measure for pairwise scaling and shifting is used. It can be adapted also to a large data set by using an adaptive wavelet-based incremental method.

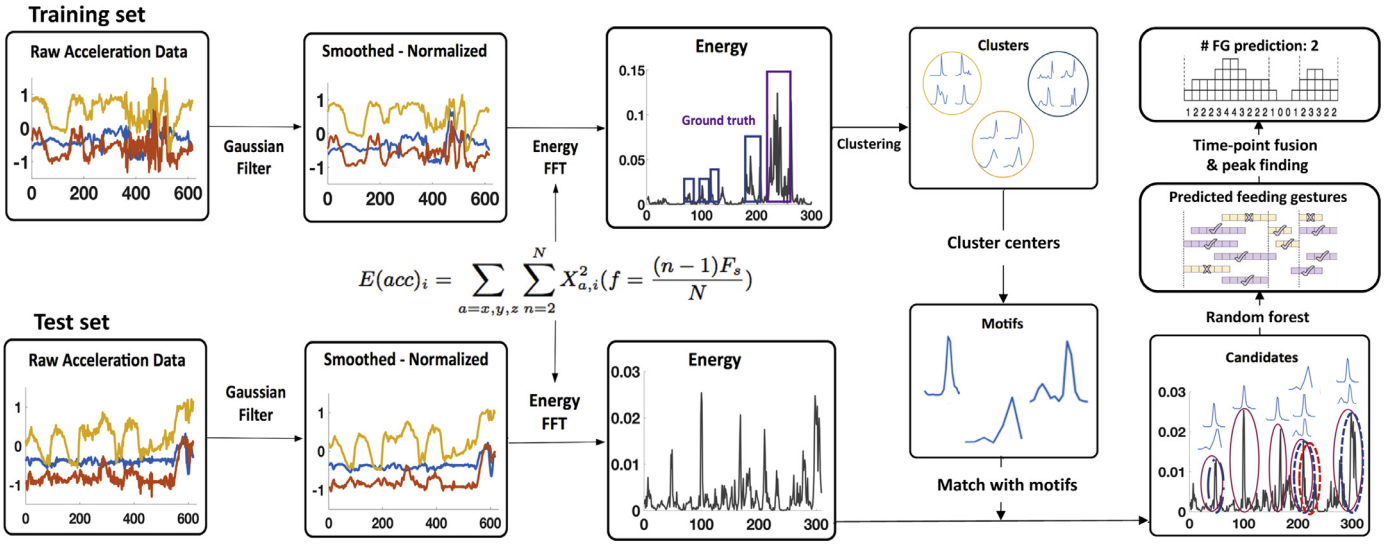

Fig. 4. An overall architecture of motif-based time-point classifier fusion classification framework.

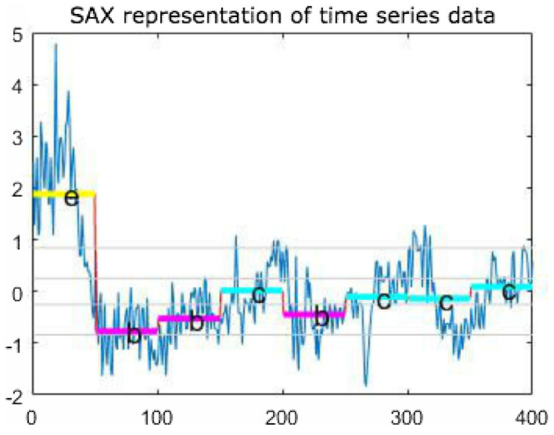

Fig. 5. Example SAX representation of 400 points raw time series data, n = 400, w = 8, a = 5.

Based on exact segmentation from the labeling, we segment the acceleration dynamic energy signal of the training set and get all the feeding gesture motifs. Then with the KSC method we obtain motif templates, the cluster centers.

4.4.3. Motif matching and segmentation based on symbolic aggregate approximation (SAX)

SAX takes two steps to transform time series data into symbolic representation. Firstly it normalizes the data and applies the Piecewise Aggregate Approximation (PAA) method. By equally slicing *n*-length time series data to *w* pieces and calculating the mean value for each segment, the *n*-length time series data is converted to *w*-length real value representation. Secondly, the *w*-length series is transformed into symbolic representation with a dictionary of alphabet size *a*. During discretization, breakpoints of the continuous series are determined by Gaussian curve because normalized time series data highly obeys Gaussian distribution. In Fig. 5, the original data and SAX representations are shown as well as the intermediate representation of eight continuous data points. With the SAX method, the 400-length time series data is translated into the string ebbcbccc.

After obtaining motif templates from KSC clustering, we apply all the motif templates on the signal to identify motif candidates. We use $MINDIST(\hat{Q}, \hat{C})$ equation to calculate the distance between the motif template and motif candidate.

$$MINDIST(\hat{Q}, \hat{C}) = \sqrt{\frac{n}{w} \sum_{i=1}^{w} (dist(\hat{q}_i - \hat{c}_i))^2}$$

$\hat{Q}$ represents the motif template, comprising word $\hat{q}_i (i = 1, \ldots, w)$. The distance between two words is found in a lookup table. In Table 7 we illustrate the lookup table with the example when alphabet size $a = 5$.

Table 7
A lookup table showing the distance between two symbols when alphabet size is 5.

	a	b	c	d	e
a	0	0	0.59	1.09	1.69
b	0	0	0	0.50	1.09
c	0.59	0	0	0	0.59
d	1.09	0.50	0	0	0
e	1.69	1.09	0.59	0	0

To reduce the numerosity problem, when two adjacent motif candidates are detected by one template, we remove the latter one. In this step, high feeding gesture recall is stressed (as opposed to high precision) since we do not want to miss any feeding gesture. We make the assumption that false positives will be filtered during the classification stage.

4.5. Feature extraction and selection

Features of data stream segments play the role of generating a fingerprint for the intended activity in machine learning and pattern recognition. Following data preprocessing of the raw signal, it is important to determine what features to collect on the raw signal that will be predictive of one's outcome. Due to the high variability across signals representing the same activity, and to ensure that the system is capable of running in real-time, 109 statistical features are extracted from each candidate feeding gesture (identified by motif matching), which is a flexible sized window. In this way, we create a fingerprint for an entire feeding gesture as opposed to part of a feeding gesture. These features, known to be useful in detecting activity [29] and eating [9], include: mean, median, max, min, standard deviation, kurtosis, interquartile range, quartile 1, quartile 3, skewness, and root mean square (RMS). With the information from the three-axis accelerometer and three-axis

Table 8
List of features. All sensor data: three-axis accelerometer data, three-axis gyroscope data, pitch and roll.

Feature	Sensor Data
Mean	All sensor data
Median	All sensor data
Max	All sensor data
Min	All sensor data
Standard	All sensor data
Deviation	All sensor data
Kurtosis	All sensor data
Interquartile range	All sensor data
Quartile 1	All sensor data
Quartile 3	All sensor data
Skewness	All sensor data
Root mean square	All sensor data
Covariance	Across different axes in accelerometer
Covariance	Across different axes in gyroscope
Covariance	Across the same axis between accelerometer and gyroscope
Covariance	Across different axes between accelerometer and gyroscope
Sum of absolute change	Pitch and roll

gyroscope, we calculate pitch and roll with following formulas [30]:

$$pitch = arctan\left(\frac{G_y}{\sqrt{G_x^2 + G_z^2}}\right)$$

$$roll = arctan\left(\frac{-G_x}{G_z}\right)$$

G_x, G_y and G_z are the three-axis accelerometer data. Running each statistical feature on each axis of the six-axis inertial sensors as well as pitch and roll generates 88 features. Calculating covariance of all the pairwise combination from the six-axis sensor data generates 15 features. Along with three-axis acceleration energy, the sum of the three-axis acceleration energy and sum of changes along pitch and roll, we created samples with 109 dimensions in the $\mathbb{R}^{109}$ feature space (Table 8).

4.6. Classification and feeding gesture count estimation

We used the Random Forest model (N = 185) built on motif segments from the training set to classify the candidate feeding gestures. As we used multiple motifs to search, the detected candidate motifs overlap with each, even after SAX numerosity reduction. Thus we predicted feeding gestures based on local maximum findings, maintaining a score for each time point in the time series (such that if a point overlapped with 10 candidate feeding gestures, it had a score of 10).

In this way, we clustered the results of motif matching and produced the final result of feeding gesture counting. By converting from segments to time point scores we obtained the score curve over time. We then looked for local maximums on this curve, restricted by the fact that two feeding gestures can not happen within two seconds (empirically found to produce best results). The number of peaks is the feeding gesture count. In this way, we grouped the candidate feeding gestures and obtained the final feeding gesture count. Fig. 6 shows how a time-point fusion and peak finding method can reveal hidden information post feeding gesture classification phase. The purple motif candidates are recognized as feeding gestures and the yellow motif candidates non-feeding.

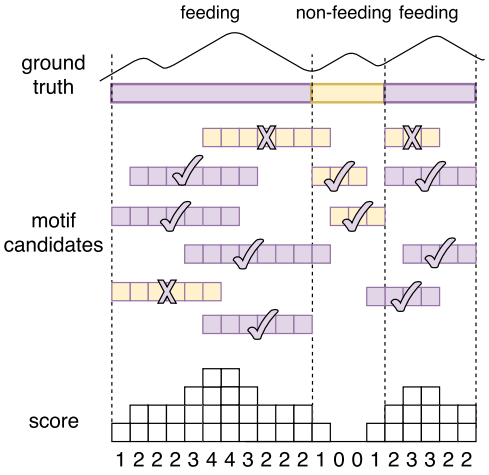

Fig. 6. Motif-based time-point fusion and peak finding method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

4.7. Defining the criterion (prediction measure)

To determine if a candidate feeding gesture was correct or not, we used event-based evaluation [31], where we defined an overlap ratio between ground truth and the candidate's feeding gesture. When the prediction covered part of the ground truth segments there were four possible scenarios: ground truth start time precedes prediction start time, prediction start time precedes ground truth start time, prediction start and end time spans ground truth, and ground truth spans the prediction. We used overlap ratio as a measurement to decide whether a segment was correctly identified or not. We defined overlap, k, as the ratio of the overlap of ground truth and prediction segments to the combination of both ground truth and prediction segments. It can be expressed in the following function, where G_1 and G_2 are separately the start time and end time of the ground truth segment, and P_1 and P_2 are separately the start time and end time of the prediction segment.

$$k = \frac{min(G_2, P_2) - max(G_1, P_1)}{max(G_2, P_2) - min(G_1, P_1)}$$

4.8. Evaluation

We focus our evaluation on eight important factors:

- Positive Predictive Value (Precision): the number of true positives (i.e. the number of segments labeled as feeding gestures) divided by the total number of segments predicted as feeding gestures by the algorithm;
- F1 score (Positive Class of Feeding Gesture): the harmonic mean of precision and recall for feeding gestures;
- TPR (True Positive Rate): the number of true positives divided by the total number of true feeding gestures, also referred to as recall and sensitivity;
- FPR (False Positive Rate): the total number of false predicted feeding gestures, divided by the total number of non-feeding gestures;
- Specificity (The True Negative Rate): the total number of correctly predicted non-feeding gestures, divided by the total number of non-feeding gestures;
- C. Kappa (Cohens Kappa): the agreement between the predicted and true labels;
- MCC (Matthews Correlation Coefficient): generally regarded as a balanced measure, even if classes have different sizes; and
- Weighted Accuracy: accuracy weighted by bias of positive and negative sample numbers.

Table 9
Performance matrics for highly Structured experiments, in-lab structured experiments and unstructured experiments with exact segmentation(averaged for 5 runs). HS, IS and US are abbreviations for highly structured experiments, in-lab structured experiments and unstructured experiments.

	Personalized (training/test set split)		Generalized (10 fold CV)		Generalized (LOPO)		Generalized training/test set split)	
Protocol	HS	IS	HS	IS	HS	IS	HS	IS
Prec(pos)	0.96	0.97	0.95	0.92	0.86	0.94	0.94	0.95
F1(pos)	0.95	0.97	0.95	0.87	0.83	0.88	0.93	0.97
TPR	0.95	0.98	0.95	0.82	0.82	0.85	0.93	0.98
FPR	0.05	0.12	0.05	0.02	0.14	0.02	0.06	0.20
Specificity	0.95	0.88	0.95	0.98	0.86	0.98	0.94	0.80
MCC	0.91	0.87	0.90	0.84	0.69	0.86	0.87	0.82
CKappa	0.91	0.87	0.90	0.84	0.68	0.85	0.87	0.82
w-acc	0.95	0.93	0.95	0.90	0.84	0.91	0.93	0.89

Our framework comprises three stages: motif matching, classification, and fusion. For each stage we have a set of metrics to assess performance. For motif matching, the goal is to discover as many feeding gesture durations as possible, without missing any feeding gestures, and so recall is our most critical measure in this step. In the classification stage, given the high overlap of candidate segments, we only want to predict a feeding gesture when we are most sure of one, and so we use precision as our main evaluation metric in this stage. Finally, after the fusion stage, accuracy and root mean square error (the difference between true feeding gesture count and predicted feeding gesture count) is our final metric to assess feeding gesture count.

4.9. Limitations

Because we focused on detecting feeding gestures based on hand-up and hand-down motion, there were at times conditions that generated unusual data, such as when the participant drank from a straw for longer than 30 s. We disregarded these segments as a single gesture.

In the in-lab structured experiment, one of the Microsoft wristbands generated inertial sensor readings at 8 Hz data for six participants, despite us setting our sampling rate at 31 Hz. As a result, when showing the results for the exact segmentation approach in the in-lab structured experiment, we down-sampled the 31 Hz data to make all ten participants show 8 Hz of data. While this perhaps only slightly affected the exact segmentation approach, it significantly affected our motif-based segmentation approach (which required higher signal granularity). As a result, we only show results for the four participants in the in-lab structured experiment using the motif-based segmentation approach. In the unstructured test, ten participants data sets were collected, but one participants feeding gestures were missing, two participants video recordings could not be synchronized due to missing snippets of video, and two participants did not eat enough (they were already full from pre-loading), and so using a personalized model for those last two participants was not possible. As a result we reported on only five of the participants in the unstructured eating experiment.

5. Results

5.1. Upper bound on feeding gestures

The results of the Random Forest classifier (number of trees n = 185) under the personalized and generalized experiments for highly structured and in-lab structured test using exact segmentation are shown in Table 9. The metrics show that the personalized model in general outperformed the generalized model. In highly structured, however, the LOPO provided the lowest precision, MCC, and highest false positive rate. The LOPO provides reasonably good results in in-lab structured test, though, suggesting that feeding gestures perhaps have more restrictive variability between subjects than pretend eating behaviors (seen in highly structured test). Most notably, these results show the importance of developing personalized models, and more importantly, obtaining accurate segmentation in order to yield positive results.

Table 10
Recall of energy based motif detection for different measurement threshold in in-lab structured test.

Criterion	0.5	0.6	0.7	0.8	0.9
Mean	0.9868	0.9231	0.7626	0.5631	0.2361
Std dev	0.0272	0.0900	0.2230	0.1953	0.1342

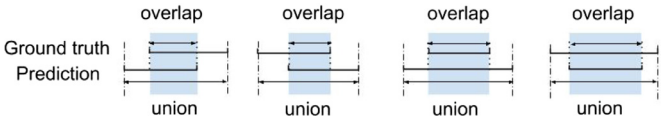

Fig. 7. A diagram of measurement criterion.

5.2. Motif generation

To generate motifs of feeding gestures, we used the K-Spectral Centroid Clustering method to generate energy-based motif templates from the training set using ground truth (exact segmentation), and then we performed motif matching using the SAX to generate candidate feeding gestures, an alphabet size $a = 5$, a motif length based on ground truth, and the number of symbols in the low dimensional approximation of the sub-sequence of $N/2$ (N is the length of time series). We set the de-noising parameter to 0.01, which means that if the standard deviation of one motif candidate was less than 0.01, then this candidate would be regarded as noise and removed. When processing signals from the test data, we generated the energy signal, normalized it, and attempted to find matches of our motif-based feeding gestures in the signal. We set the distance threshold as 0.7, so if the distance between a true motif and a candidate segment is larger than 0.7, the candidate feeding gesture is not matched. This approach generates several candidate feeding gestures with a high recall as shown in Table 10, based on the criterion defined in Section 4.6.1 and dummyTXdummy- in Fig. 7. The smaller the criterion, the less required overlap between ground truth and the predicted signal, and the higher the recall. We selected a loose setting of the criterion to achieve high recall ($k = 50\%$), so that we did not miss any feeding gestures. We relied on the classifier to distinguish the true segments.

We also tested motif-based segmentation based on single axes of acceleration, and Table 11 provides results when using the x-axis of acceleration. We show significant improvement in recall when

Table 11
Recall of acceleration based motif detection for different measurement threshold in in-lab structured test.

Criterion	0.5	0.6	0.7	0.8	0.9
Mean	0.9193	0.8101	0.6245	0.4756	0.1509
Std dev	0.0855	0.1192	0.2007	0.2231	0.0861

using the energy-based motif segmentation methodology as shown in Table 10, and maintained energy as the signal to represent feeding gesture motifs.

5.3. Classification results based on motif generation

We split the in-lab structured data segments into two sets (for each activity collected): 70% of the segments were made part of the training set and 30% of segments were set aside for the test set. Using the ground truth labels of the training set and the features, we designed a Random Forest model with the number of trees set to 185, then applied the model on the test set. Table 12 provides results for the in-lab structured data on the four participants with 31 Hz. The results for motif-based segmentation did not perform well for the participants with 8 Hz data, showing the importance of granularity in motif-based segmentation.

To prove the effectiveness of a motif-based segmentation method, we compared it against the fixed-length sliding window segmentation methodology tested by Edison et al. in [9]. We set the segmentation window size to 2 s, and the stride to 1 s. After extracting the features defined in prior literature to be effective in capturing feeding gestures, we applied the Random Forest classifier and achieved the results shown in Table 12, yielding an average F-score of 68%. On the other hand our motif-based segmentation algorithm showed a 7% improvement resulting in an average 75% F-score. We tried to use the data provided by Edison et al., but, found that there were no labels for the start and end of a feeding gesture, but rather for bites, which would not be sufficient for feeding gesture count.

5.4. Feeding gesture estimation

We used the Random Forest model built on training set motif segments to classify the candidate feeding gestures in the test set. We discovered that several classified feeding gestures overlapped with a single feeding gesture, since motif-based segmentation resulted in several overlapping candidate feeding gestures. As a result, for every time-series sample point we counted the number of candidate feeding gestures that overlapped with that time point, and provided a score for each time point (a sum of the overlapped feeding gestures), as shown in Fig. 8(b). The top Fig. 8(a) shows the feeding gesture ground truth: 0 means no feeding gesture, 1 means feeding gesture. The figure in the middle is the score for each time point, voted based on the segment classification result. The number of peaks are shown in red circles. The bottom figure shows the feeding gesture prediction result. We also performed post-processing to filter any feeding gestures within two seconds of one gesture. This helped mitigate the effect of noise.

With the method described in Section 4.5, we did experiments on the unstructured data set and obtained an average accuracy of feeding gesture count of 95.0%.

Because there are several candidate feeding gestures that overlap with every true feeding gesture, precision of the classification algorithm as a measure is more important than recall. In Table 13, the average precision is 87%, and the average feeding gesture accuracy is 94%, resulting in a root mean square error (RMSE) of 2.9.

5.5. Overeating prediction

The goal of the unstructured study was to distinguish overeating from non-overeating. We calculated the average calorie per feeding gesture by averaging the calorie per feeding gesture across all participants (12.39 kiloCalories). After generating the feeding gesture count for each participant, their corresponding caloric intake was calculated as shown in Table 14. While caloric intake prediction is not accurate, we accurately predicted whether a participant overate or not based on our 50% Harris Benedict definition. We believe that more accurate caloric intake prediction could be obtained if a personalized calorie per feeding gesture count is used for each participant.

6. Discussion and future work

In this work we collected data from a series of experiments following three protocols. From the correctly collected data we could predict if a participant was overeating based on the Harris Benedict principle combined with sensor data.

From this work we gained empirical knowledge that the 8 Hz sensor does not support effective feeding gesture recognition compared to using 31 Hz sensor data. Admittedly, with low frequency data the power of motif-based searching method is weakened because the detailed description of the feeding gesture is lost due to fewer points in the motifs generated (especially given the short duration of a feeding gesture).

When generating the motifs, we used a one-dimensional time series motif discovery algorithm. In the future we could try a multi-dimensional time series motif discovery algorithm. As the quality of motif decides the effectiveness of segmentation, further study for different motif generation methods is worthwhile. This future research would allow for utilizing data from both accelerometer and gyroscope sensors, or utilizing three-axis sensor data directly instead of generating a single energy-based signal, where information from each axis is lost.

While our technique outperforms sliding window, there remains the advantage of using the sliding window technique to develop real-time algorithms. In our effort, however, we are attempting to understand problematic eating behaviors and overeating, paving the way for more work on real-time systems in the near future.

Table 12
Motif based segmentation result in In-lab Structured (IS) test, RMSE = 9.5.

		Sliding Window	Motif-based Time-point Fusion Classification (MTFC)				
Subject	# FG	Classification F-score(pos)	Motif Recall	Classification F-score(pos)	Classification Precision(pos)	# FG Prediction	FG Accuracy
P3	36	0.62	0.94	0.77	0.93	32	0.89
P4	59	0.72	0.8	0.78	0.86	44	0.75
P5	42	0.71	1	0.79	0.93	42	1
P10	39	0.67	0.79	0.72	0.82	28	0.72
Average	44	0.68	0.88	0.75	0.89	36.5	0.84

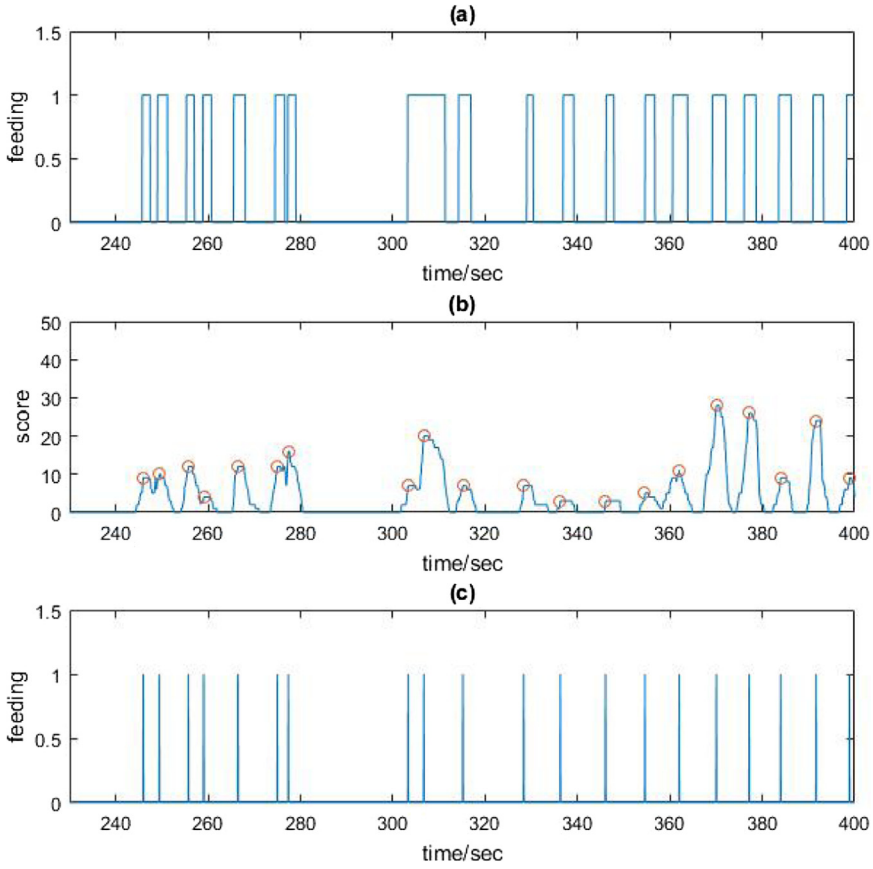

Fig. 8. Feeding gesture count result for 8th subject: (a) shows the ground truth feeding gesture duration, (b) shows the score for each time point from prediction, red circle represents peak, (c) shows the start points of predicted feeding gestures from (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

Table 13
Feeding gesture prediction result of test set in UnStructured (US) test, RMSE = 2.9.

Subject	# FG ground truth	Classification Precision(pos)	Classification F-score(pos)	# FG prediction	# FG accuracy
P1	15	0.70	0.36	13	0.87
P3	26	0.97	0.16	26	1
P6	34	0.88	0.43	35	0.97
P7	51	0.90	0.50	45	0.88
P8	70	0.88	0.07	69	0.99
Average	39.2	0.87	0.30	37.6	0.94

Table 14
Feeding gesture prediction result of test set in unstructured test.

Subject	Calories prediction	Real intake calories	50% Harris Benedict	Overeat prediction
P1	571	180	823	Not overeat
P3	1081	1642	971	Overeat
P6	1462	1226	1088	Overeat
P7	1937	1813	971	Overeat
P8	2662	1408	1051	Overeat

Due to the known numerosity problem in using the SAX method, neighboring subsequences extracted are often similar to each other, resulting in a large number of segments returned as candidate feeding gestures (high recall, but low precision). In this effort, we reduce the numerosity by applying numerosity reduction, keeping only one similar sequence from two adjacent segments. In the future, a hierarchical clustering technique may be deployed to further reduce the size of the segments.

Generating high quality data is essential to building accurate models of detection, and takes a non-negligible role in developing an effective feeding gesture recognition study. In further work, more meals could be recorded for one subject which may allow researchers to build better models and have more evidence to justify the advantage of a proposed model or framework.

In the future we aim to embed the model with the wearable wrist-worn sensor using efficient sensing [32], flexible in-node processing, and rapid prototyping, similar to the SPINE framework defined by Andreoli et al. [33].

7. Conclusion

In this effort we prove that we can predict overeating through feeding gesture counts and the Harris Benedict principle. We employ a novel motif-based time point classification framework (MTC). The result shows that the new direction of the motif-based segmentation can receive better results than fixed-length sliding window segmentation. By showing the upper bound for classifica-

tion between feeding/non-feeding gestures using ground truth exact segmentation, we showed that if we are able to solve the automated segmentation problem accurately, we can achieve high feeding gesture count accuracy. We showed that we are able to obtain an average RMSE of 9.5 for in-lab structured eating, and an even better RMSE of 2.0 in unstructured eating. Thus more study along this path is worth further exploration. With classification and a score for peak method, the feeding gesture count could be predicted with average 94% accuracy, and the overeating period could be identified for five subjects in an unstructured data set. These results thus show promise and pave the way for detection of overeating in populations with problematic eating behaviors.

References

[1] A. Goldschmidt, M. Jones, J. Manwaring, The clinical significance of loss of control over eating in overweight adolescents, Int. J. Eat. Disord. 41 (2) (2008) 153–158.

[2] C.P. Herman, J. Polivy, External cues in the control of food intake in humans: the sensory-normative distinction, Physiol. Behav. 94 (5) (2008) 722–728.

[3] C.G. Greeno, R.R. Wing, Stress-induced Eat. 115 (3) (1994) 444–464.

[4] J. Cools, D.E. Schotte, R.J. McNally, Emotional Arousal Overeating Restrained Eaters. 101 (2) (1992) 348–351.

[5] P. Bongers, A. Jansen, R. Havermans, A. Roefs, C. Nederkoorn, Happy eating: the underestimated role of overeating in a positive mood, Appetite 67 (2013) 74–80.

[6] O.D. Lara, M.A. Labrador, A survey on human activity recognition using wearable sensors, Commun. Surv. Tutorials, IEEE 15 (3) (2013) 1192–1209.

[7] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80, doi:10.1016/j.inffus.2016.09.005.

[8] H. Junker, O. Amft, P. Lukowicz, G. Tröster, Gesture spotting with body-worn inertial sensors to detect user activities, Pattern Recognit. 41 (6) (2008) 2010–2024, doi:10.1016/j.patcog.2007.11.016.

[9] E. Thomaz, I. Essa, G.D. Abowd, A practical approach for recognizing eating moments with wrist-mounted inertial sensing, in: Proceedings of the 2015 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing, in: UbiComp '15, ACM, New York, NY, USA, 2015, pp. 1029–1040.

[10] Y. Dong, J. Scisco, M. Wilson, E. Muth, A. Hoover, Detecting periods of eating during free-living by tracking wrist motion, IEEE J. Biomed. Health Inform. 18 (4) (2014) 1253–1260, doi:10.1109/JBHI.2013.2282471.

[11] O. Amft, H. Junker, G. Troster, Detection of eating and drinking arm gestures using inertial body-worn sensors, in: Proceedings of the Ninth IEEE International Symposium on Wearable Computers, in: ISWC '05, IEEE Computer Society, Washington, DC, USA, 2005, pp. 160–163, doi:10.1109/ISWC.2005.17.

[12] T. Rahman, M. Czerwinski, R. Gilad-Bachrach, P. Johns, Predicting "about-to-eat" moments for just-in-time eating intervention, in: Proceedings of the 6th International Conference on Digital Health Conference, in: DH '16, ACM, New York, NY, USA, 2016, pp. 141–150, doi:10.1145/2896338.2896359.

[13] C. Maramis, V. Kilintzis, N. Maglaveras, Real-time bite detection from smartwatch orientation sensor data, in: Proceedings of the 9th Hellenic Conference on Artificial Intelligence, in: SETN '16, ACM, New York, NY, USA, 2016, pp. 30:1–30:4, doi:10.1145/2903220.2903239.

[14] O. Amft, G. Tröster, On-body sensing solutions for automatic dietary monitoring, IEEE Pervasive Comput. 8 (2) (2009) 62–70.

[15] S. Zhang, R. Alharbi, W. Stogin, K. Moran, A.F. Pfammatter, B. Spring, N. Alshurafa, Machine learning algorithms applied to detect feeding gestures, abstract presentation at: The Obesity Society Annual Meeting at ObesityWeekSM 2016, New Orleans, LA, 2016, www.obesityweek.com.

[16] S. Sen, V. Subbaraju, A. Misra, R.K. Balan, Y. Lee, The case for smartwatch-based diet monitoring, in: 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), 2015, pp. 585–590, doi:10.1109/PERCOMW.2015.7134103.

[17] N. Saleheen, A.A. Ali, S.M. Hossain, H. Sarker, S. Chatterjee, B. Marlin, E. Ertin, M. al'Absi, S. Kumar, puffmarker: a multi-sensor approach for pinpointing the timing of first lapse in smoking cessation, in: Proc. of the 2015 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing, in: UbiComp '15, ACM, New York, NY, USA, 2015, pp. 999–1010.

[18] E. Sazonov, P. Lopez-Meyer, S. Tiffany, A wearable sensor system for monitoring cigarette smoking, J. Stud. Alcohol Drugs 74 (6) (2013) 956–964.

[19] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, E. Kalogerakis, Risq: recognizing smoking gestures with inertial sensors on a wristband, in: Proceedings of the 12th Annual Int. Conf. on Mobile Systems, Applications, and Services, in: MobiSys '14, ACM, New York, NY, USA, 2014, pp. 149–161.

[20] Y. Dong, A. Hoover, J. Scisco, E. Muth, A new method for measuring meal intake in humans via automated wrist motion tracking, Appl. Psychophysiol Biofeedback 37 (3) (2012) 205–215.

[21] Y. Dong, J. Scisco, M. Wilson, E. Muth, A. Hoover, Detecting periods of eating during free-living by tracking wrist motion, IEEE J. Biomed. Health Inform. 18 (4) (2014) 1253–1260.

[22] N. Saleheen, A.A. Ali, S.M. Hossain, H. Sarker, S. Chatterjee, B. Marlin, E. Ertin, M. al'Absi, S. Kumar, puffmarker: a multi-sensor approach for pinpointing the timing of first lapse in smoking cessation, in: Proceedings of the 2015 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing, in: UbiComp '15, ACM, New York, NY, USA, 2015, pp. 999–1010.

[23] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing sax: a novel symbolic representation of time series, Data Mining Knowl. Discov. 15 (2) (2007) 107–144, doi:10.1007/s10618-007-0064-z.

[24] A. Fouse, N. Weibel, E. Hutchins, J.D. Hollan, Chronoviz: a system for supporting navigation of time-coded data, in: CHI '11 Extended Abstracts on Human Factors in Computing Systems, in: CHI EA '11, ACM, New York, NY, USA, 2011, pp. 299–304, doi:10.1145/1979742.1979706.

[25] S.H. Roza AM, The harris benedict equation reevaluated: resting energy requirements and the body cell mass, Am. J. Clin. Nutr. 40 (1) (1984).

[26] S. Zhang, R. Alharbi, M. Pourhomayoun, W. Stogin, N. Alshurafa, Food watch: detecting and characterizing eating episodes through feeding gestures, in: International Conference on Body Area Networks, BodyNets 2016, Italy, 2016.

[27] A.V. Oppenheim, A.S. Willsky, S.H. Nawab, Signals & systems (Second ed.), Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[28] J. Yang, J. Leskovec, Patterns of temporal variation in online media, in: ACM International Conference on Web Search and Data Minig (WSDM), Stanford InfoLab.

[29] N. Alshurafa, W. Xu, J.J. Liu, M.C. Huang, B. Mortazavi, M. Sarrafzadeh, C. Roberts, Robust human intensity-varying activity recognition using stochastic approximation in wearable sensors, in: 2013 IEEE Int. Conf. on Body Sensor Networks, 2013, pp. 1–6.

[30] M. Pedley, Tilt sensing using a three-axis accelerometer, Freescale Semiconductor Application Note (2013), 2012–2013.

[31] M. Everingham, J. Winn, The pascal visual object classes challenge (voc2007) development kit, Technical Report, 2007.

[32] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, R. Jafari, Enabling effective programming and flexible management of efficient body sensor network applications, IEEE Trans. Hum. Mach. Syst. 43 (1) (2013) 115–133, doi:10.1109/TSMCC.2012.2215852.

[33] A. Andreoli, R. Gravina, R. Giannantonio, P. Pierleoni, G. Fortino, SPINE-HRV: A BSN-Based Toolkit for Heart Rate Variability Analysis in the Time-Domain, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 369–389.