IDEA: Instant Detection of Eating Action using Wrist-Worn Sensors in Absence of User-Specific Model*

Junghyo Lee, Prajwal Paudyal, Ayan Banerjee, Sandeep K.S. Gupta
iMPACT Lab (http://impact.asu.edu), CIDSE, Arizona State University
{jlee375,ppaudyal,abanerj3,sandeep.gupta}@asu.edu

ABSTRACT

Eating activity monitoring using wearable sensors can potentially enable interventions based on eating speed for critical healthcare problems such as obesity or diabetes. We propose a novel methodology, IDEA that performs accurate eating action identification and provides feedback on eating speed. IDEA uses a single wristband with IMU sensors and functions without any manual intervention from the user. The F1 score for eating action identification was 0.92.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**;

KEYWORDS

Wearable; Gesture; User Adaptive Modeling; Diet Monitoring

ACM Reference Format:
Junghyo Lee, Prajwal Paudyal, Ayan Banerjee, Sandeep K.S. Gupta. 2018. IDEA: Instant Detection of Eating Action using Wrist-Worn Sensors in Absence of User-Specific Model. In *UMAP'18: 26th Conference on User Modeling, Adaptation and Personalization, July 8-11, 2018, Singapore.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3209219.3209265

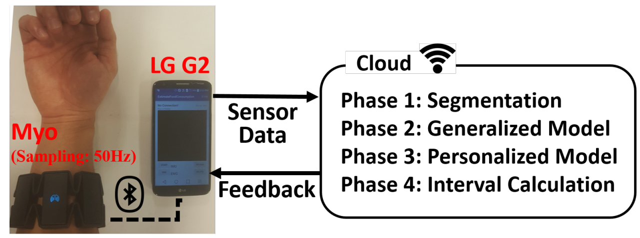

Figure 1: IDEA architecture and overview.

1 INTRODUCTION

Eating action monitoring is important for administering and facilitating eating speed based dietary interventions. In the paper, we propose *IDEA*, Instant Detection of Eating Action that can operate with any wristband based IMU sensors (accelerometers, orientation, and gyroscope) to instantly identify eating action without any manual input from user.

Challenge: An eating action is a sequential arrangement of three distinct components interspersed with gestures that may be unrelated to the eating action. This makes it extremely challenging to accurately identify eating actions. The primary reason for the lack of acceptance of state-of-art eating action monitoring techniques [3] includes: i) the need to install wearable sensors that are cumbersome to wear or limit mobility of the user, ii) the need for manual user input, and iii) poor accuracy in absence of adequate user input.

The core hypothesis of *IDEA* is that despite variations in arm movements, locations of the mouth and food plate, type of food, and utensils used, an eating action is universal and is expected to have commonalities amongst individuals. For example, if a person is using an utensil to pick up food from a plate on a table, the arm action to lift the food up from the table to mouth is assumed to be common across all individuals. Once such common actions are identified, other information such as picking action, putting food in mouth, twisting of wrist to orient spoon or fork towards the mouth can then be added to the common action to identify whether that common action is a part of an eating action. To exploit such commonalities, *IDEA* uses a two-step process, **i) Generalized Model** and **ii) Personalized Model**. Also, *IDEA* uses a small set of users, **donors**, from which the training data is collected to derive a set of eating action candidates. A practical deployment of *IDEA* does not need any data from a given user and identifies eating actions in a plug-n-play manner. *IDEA* in effect provides automated labeling.

2 METHODOLOGY

The *IDEA* methodology consists of four phases as seen Fig. 1: segmentation, generalized model, personalized model, and interval calculation. For segmentation, we utilize an extrema based segmentation method. From our observation of dataset, we conclude that when the user starts and finishes any eating action component, their hand pauses momentarily or there is a sharp change in the orientation of their wrist. Based on this observation, we utilized the extrema to segment the continuous movement of hand gesture into two types of segments: a) "Eat" and b) "No Eat". The extrema based segmentation method generates irregular size segments so we used the interpolation to obtain uniform size segments. The aim of generalized model, where data from an individual A is compared with the data of other users in a set S, is to detect strong and weak candidates for eating action of A. This is done using a Deep Neural Networks (DNN), four hidden layers with nodes starting from 512 and exponentially reducing to 64. The activation function is ReLU, the output layer is sigmoid for binary classification. The output of this phase is two folds: i) two sets including the set of actions that are confirmed eating actions, and unconfirmed ones, and ii) a set of users in the set S that are "similar" to the given user. In the

*Partly funded by NIH NIBIB (Grant #EB019202) and NSF IIS (Grant #1116385).

ACM ISBN 978-1-4503-5589-6/18/07.
https://doi.org/10.1145/3209219.3209265

personalized model phase, the eating actions obtained from the set of "similar" users given by generalized model are used as training set to classify the unconfirmed actions set as true eating actions or not. The interval calculation provides the eating speed feedback. In this phase, we obtain eating intervals by simply calculating the time-stamp difference between previous 'Eat' and current 'Eat'.

3 RELATED WORK

Nearly all eating behavior monitoring systems require an initial training phase, where the user must provide labeled data related to an eating action as in Sen et al. [6]. This initiation task is time consuming and often annoying. Moreover, such training must be redone if the food item, plate, and utensil changes. *IDEA* is plug-n-play and automates the training process by first using a general model to detect few eating actions and using them as training data.

A distracted eating pattern is when a user is involved in other activities like talking, swallowing saliva, shifting in their seats, and multiple picking before eating. Liu et al. [4] could not account for distracted eating situations. However, since *IDEA* identifies the three actions separately, it can detect such distracted situations.

4 EXPERIMENTAL RESULT

Setup: Fig. 1 displays *IDEA* architecture. The user wears a Myo wristband [1] which collects accelerometer, orientation, gyroscope, Electromyogram (EMG) data at a frequency of 50Hz. LG G2 (smartphone) is connected to the Myo through Bluetooth to receive Myo data. For each user, we also recorded video simultaneously using LG G2 camera. The video data is used to build the ground truth. Myo provides 18 data streams from four sensors including 3 accelerometer data streams, 4 orientation data streams, 3 gyroscope data streams, and 8 EMG data streams.

Data Acquisition: We recruited thirty-six subjects following IRB approvals. Each subject participated as a volunteer for an eating episode that lasted for at least 15 mins with an average of 30 eating actions. Subjects were asked to sit facing a smartphone camera during the eating episode and wore the Myo. Subjects ate food either obtained from a restaurant or cooked at home in different types of containers. The subjects brought their own containers and used two types of utensils: a fork and spoon. They were free to eat whatever they want. Hence, the collected data has user-dependent eating factors resulting in variations on eating directions, speed, relative distance between the food plate and the mouth, etc. In the collected data, each subject has at least 20 eating actions, which are distributed across different areas in the food plate. Also, there are a total of 1246 eating actions and 8400 non-eating gestures.

Labeling: It is important to develop the ground truth. After synchronization between the video and Myo data, we annotated the meal video manually through visual inspection. To reduce human error the annotations were performed by four independent human observers and were cross validated against each other. Each annotated video is labeled as one of the following: (1) picking, (2) carrying, and (3) putting in mouth. A cycle of these three annotated eating components is considered as one eating action. Then, the sensor data streams are labeled based on the annotated videos.

Result: With a training set of eight donors and a test set of 28, the precision of eating action identification was 0.93 while the recall was 0.89. For the worst case users, on an average, *IDEA* improves precision by 0.11 and recall by 0.15 with respect to other deep learning strategies without getting any training data or any manual user interventions. *IDEA* can also be used for automated labeling of eating action. The mislabeling rate for *IDEA* is 11 out of nearly 10,000 eating or non-eating actions and that for human eye is 18 as observed in our study. Fig. 2 shows that on an average *IDEA* improves the F1 score by 0.05 for 8 donors with respect to the DNN based approaches. *IDEA* has an F1 score of more than 0.9 if at least 7 users are included in the training set. The figure also shows the performance for the worst case user set. On an average *IDEA* improves F1 score by 0.15 for the worst case users with respect to traditional DNN based approach.

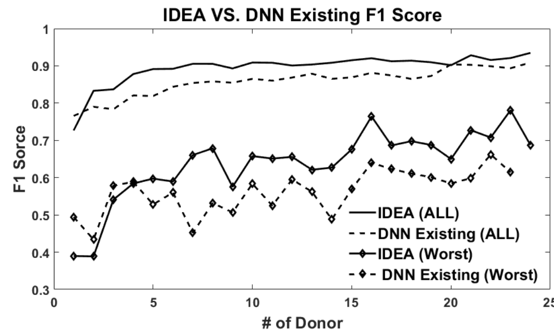

Figure 2: F1 Score Comparison between IDEA and Existing DNN. (ALL) is for all users and (Worst) is for worst users.

5 CONCLUSIONS

In this paper, we have proposed *IDEA*, a novel methodology for detecting eating actions using only a wristband sensor without the need for collecting training data from the user. The proposed methodology is plug-n-play and does not need any initialization from the user, hence working in an user-independent manner. *IDEA* can also be used to automatically annotate eating actions for future use in a personalized model. When combined with image based food type identification projects such as MT-Diet [2], *IDEA* can be applied to build a nutritional retrieval system. Also, the *IDEA* methodology will be useful for wristband based sign language recognition projects such as DyFAV [5]. Plug-n-play recognition of such complex gestures can result in fast and accurate sign language translation systems and will be explored as a crucial future work.

REFERENCES

[1] Myo Armband. https://www.myo.com Accessed: 2015-07-10.
[2] Junghyo Lee, Ayan Banerjee, and Sandeep KS Gupta. 2016. MT-diet: Automated smartphone based diet assessment with infrared images. In *PerCom*. IEEE, 1–6.
[3] Junghyo Lee, Prajwal Paudyal, Ayan Banerjee, and Sandeep KS Gupta. 2017. FIT-EVE&ADAM: Estimation of Velocity & Energy for Automated Diet Activity Monitoring. In *Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on*. IEEE, 1071–1074.
[4] Jing Liu, Edward Johns, Louis Atallah, Claire Pettitt, Benny Lo, Gary Frost, and Guang-Zhong Yang. 2012. An intelligent food-intake monitoring system using wearable sensors. In *Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on*. IEEE, 154–160.
[5] Prajwal Paudyal, Junghyo Lee, Ayan Banerjee, and Sandeep KS Gupta. 2017. DyFAV: Dynamic Feature Selection and Voting for real-time recognition of fingerspelled alphabet using wearables. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 457–467.
[6] Sougata Sen, Vigneshwaran Subbaraju, Archan Misra, Rajesh Krishna Balan, and Youngki Lee. 2015. The case for smartwatch-based diet monitoring. In *Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 585–590.