Personalised meal eating behaviour analysis via semi-supervised learning

Alexandros Papadopoulos[1], Konstantinos Kyritsis[1], Ioannis Sarafis[1] and Anastasios Delopoulos[1]

***Abstract*— Automated monitoring and analysis of eating behaviour patterns, i.e., "how one eats", has recently received much attention by the research community, owing to the association of eating patterns with health-related problems and especially obesity and its comorbidities. In this work, we introduce an improved method for meal micro-structure analysis. Stepping on a previous methodology of ours that combines feature extraction, SVM micro-movement classification and LSTM sequence modelling, we propose a method to adapt a pre-trained IMU-based food intake cycle detection model to a new subject, with the purpose of improving model performance for that subject. We split model training into two stages. First, the model is trained using standard supervised learning techniques. Then, an adaptation step is performed, where the model is fine-tuned on unlabeled samples of the target subject via semi-supervised learning. Evaluation is performed on a publicly available dataset that was originally created and used in [1] and has been extended here to demonstrate the effect of the semi-supervised approach, where the proposed method improves over the baseline method.**

I. Introduction

Obesity has been characterized as a modern epidemic, with almost a quarter of the planet's population being affected. Contrary to other epidemics, obesity can be viewed as the result of a number of factors, most of which are highly treatable, thus rendering the disease itself preventable. One such factor has been identified in the eating behavior patterns of individuals [2]. Therefore, automated and unobtrusive monitoring of the in-meal eating behavior is a research direction that can provide substantial benefits to the study and treatment of obesity [3].

Pre-existing works attempted to measure meal eating behavior using specialized equipment such as a weight scale [4] or intra ear microphones [5], while others resort to more readily available devices such as *Inertial Measurement Unit (IMU)* sensors (e.g [6]) with results highlighting the potential of using general-purpose devices, such as wristbands or smartwatches, to achieve good performance on the task.

Based on these early results, the authors of [1] recently proposed an IMU-based aproach that models the food intake cycles based on distinct movements associated with food intake and referred to as *micro-movements* (see TABLE I). They use an array of binary *Support Vector Machine (SVM)* classifiers to model the different micro-movements, followed by two *Hidden Markov Models (HMM)* in order to capture the time dependence between them. This method has led to promising results. The later work of [7] refined the original approach by substituting the HMM component with a *Long Short Term Memory (LSTM)* [8] network, leading to state-of-the-art performance in the publicly available *Food Intake Cycle (FIC)* dataset.

A potential pitfall of existing approaches is that model training occurs in a subject-agnostic way; that is, variances in the eating styles of different subjects are not taken into account. Hence, when a pre-trained model is deployed for use by a new subject, the inability to adapt it to that subject's eating style may cause performance to suffer. This could be the case when the eating style of the new subject differs from those the model has seen during training.

In a similar context, the authors of [9] recently proposed a chewing detection model that can adapt to a new user through an active learning framework that asks the user to provide feedback on a few of the model's less confident predictions. The model is then re-trained using the initial training set, augmented with the user feedback and the process is iteratively repeated until satisfactory performance is obtained.

In this work, we tackle the problem of adapting an IMU-based food intake detection model to a new subject, through the use of meal session recordings of that subject for which the ground truth is not available. This setup corresponds to a *semi-supervised learning* [10] task. Semi-supervised learning lies at the intersection of supervised and unsupervised learning. It can be seen as a learning situation where, in addition to a set of observations $X_l = \{\mathbf{x}_1, \ldots, \mathbf{x}_l\}$ and their corresponding labels $Y = \{y_1, \ldots, y_l\}$, a set of unlabeled observations $X_{ul} = \{\mathbf{x}_{l+1}, \ldots \mathbf{x}_{l+u}\}$ is also available and the goal is to exploit it, in order to derive a better classification rule.

We propose to split model training into two distinct stages: First, the LSTM model is trained in a supervised way as in [7]. Then, a second, finetuning step of the LSTM component is performed, by using an additional semi-supervised loss for the unlabeled meal sessions of the new subject. The overall method is described in detail in Section II.

Training and evaluation was carried out in an extension of the FIC dataset, where additional meal sessions were recorded in order to serve as unlabeled samples for our algorithm. Early results indicate the validity of the proposed method.

II. Proposed approach

Unlabeled samples can be useful to the learning process given that certain assumptions regarding the input distribution hold. One common such assumption is the *low-density separation* [10] of classes, which states that high-density regions in the input space, corresponding to different classes,

[1]Multimedia Understanding Group, Information Processing Laboratory, Dept. of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece

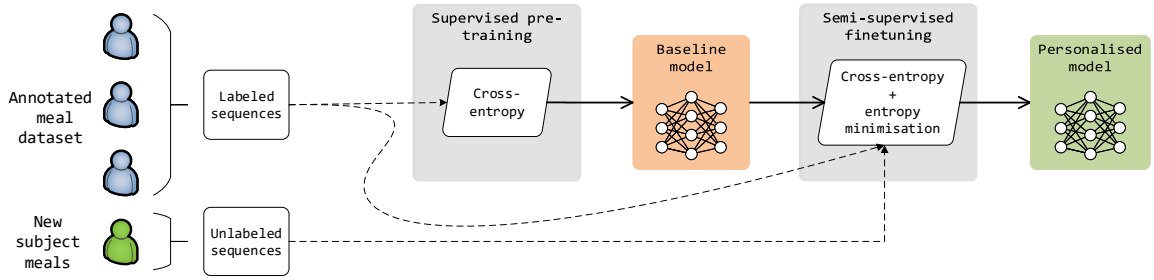

Fig. 1: Overview of the proposed method. The final intake cycle detection model is personalized for the new subject based on unlabeled sequences extracted from recorded meal sessions.

should be separated by low-density regions where samples are unlikely to be observed.

One way to enforce this assumption during training of a classifier is to add an additional term to its loss function, that causes the decision boundary to move away from dense regions. The *entropy minimisation* framework introduced in [11], achieves this effect by adding a regularisation term to the loss of a K-class probabilistic classifier, that corresponds to the conditional entropy of the classifier output distribution p_{model}, over the unlabeled samples:

$$L_{\text{ent}} = - \mathop{\mathbb{E}}_{\mathbf{x} \in X_{ul}} \left[\sum_{k=1}^{K} p_{\text{model}}(y = k|\mathbf{x}) \cdot log\ p_{\text{model}}(y = k|\mathbf{x}) \right]$$

We incorporate this idea to the framework introduced in [7], in an effort to perform semi-supervised adaptation of a pre-trained food intake cycle detection model to a new subject. We use the same model architecture and split the learning procedure into two stages:

1) Learning the base LSTM model in a supervised fashion.
2) Finetuning the base model using the entropy minimisation framework.

The rest of this section describes the two training stages in detail.

A. Supervised pre-training

The work of [7] defined a *food intake cycle* as a sequence of *micro-movements*. Micro-movements constitute a set of distinct hand movements commonly occuring during food intake cycles. The authors identified 6 distinct micro-movements that are characteristic of food intake cycles, such as movement of hand to and from plate, insertion of food into mouth, etc. The various micro-movements are listed in TABLE I.

Let $(a_x[n], a_y[n], a_z[n])$ and $(g_x[n], g_y[n], g_z[n])$ denote respectively the accelerometer and gyroscope streams captured during a meal session. Given the two streams, a data pre-processing step was performed to smooth the two signals and remove the graviational component from the acceleration signal. This was followed by a feature extraction step, where a set of both time and frequency domain features were extracted using a sliding window approach with a window size w_l of 0.2s and a step w_s of 0.1s.

Each extracted feature vector $\mathbf{f}_i$, was associated with the micro-movement the subject was performing at the respective time frame. Using this information, an array of one-vs-one SVM classifiers with RBF kernel was trained and used to convert each $\mathbf{f}_i$ into a N-dimensional SVM score vector, with $N = 10$ due to the use of 5 (the micro-movement O was not modeled due to its high-variance) micro-movements and the one-vs-one nature of the SVM scheme.

The process so far, has transformed the input signals into a sequence $\tilde{\mathbf{S}}$ of SVM score vectors $\mathbf{s}_i \in R^N, i = 1, 2, \ldots, n$, where n is the number of windows extracted from the raw signal. A sub-sequence $\mathbf{S} \subset \tilde{\mathbf{S}}$ was assigned to the positive class (intake cycle) if its corresponding micro-movement ground-truth starts with P, ends with D and contains at least one M micro-movement. The remaining sub-sequences that appear between consecutive intake cycles were associated with the negative class (non-intake cycles). The pairs of sub-sequences and labels created this way were used to train an LSTM network to classify new sequences of $\mathbf{s}_i$ as intake or non-intake cycles. To this end, a two layer LSTM with 128 cells in each layer was employed, followed by a fully connected layer with one output unit that provided the probability of intake cycle. The network was trained using the cross-entropy loss coupled with the RMSprop[1] optimizer.

We use the above architecture, but perform some changes to the LSTM component. In particular, we apply dropout to both the input and the recurrent connections of the LSTM and add an extra unit to the dense output layer, so that it explictly provides the probabilities of both the positive (intake cycle) and the negative (non-intake cycle) class. This results in an overparameterised version of the model distribution. Both the original and the overparameterised versions can describe the same set of probability distributions, but the latter allows computing the semi-supervised loss of the next section in a numerically stable way.

Throughout this work, we will assume that the SVM part of the architecture is fixed and serves as a feature extraction module. Thus, in the following, the term "model" will always refer to the LSTM network.

B. Semi-supervised finetuning

After the model has been trained in a strictly supervised manner for E_1 epochs, we perform a finetuning step by

[1] https://goo.gl/bcLNH4

Micro-movement	Description
Pick food	Hand manipulates a utensil to pick food from the plate
Upwards	Hand moves upwards, towards the mouth area
Downwards	Hand moves downwards, away from the mouth area
Mouth	Hand inserts food in mouth
No movement	Hand exhibits no movement
Other movement	Every other hand movement

TABLE I: Listing of all identified micro-movements

minimising the class conditional entropy of the model over unlabeled $\mathbf{s}_i$ sequences extracted from additional meal intake sessions. By doing so, we are implicitly asking the model to alter the decision boundaries it has learnt thus far, so that they do not pass from high-density regions of samples from the subject we are adapting to. This has the potential of correcting the decision boundary in regions of the input space that were previously unexplored due to the lack of representative training samples.

More specifically, let $S_{ul} = \{\mathbf{S}_1, \mathbf{S}_2, \dots \mathbf{S}_n\}$ be a set of unlabeled sequences of $\mathbf{s}_i$ vectors for a particular subject. We assume that each $\mathbf{S}_j \in S_{ul}$ is associated with exactly one class, although we do not know which one. In addition, let $S_l = \{\mathbf{S}_1, \mathbf{S}_2, \dots \mathbf{S}_m\}$ be the set of labeled sequences the model has already been trained on. The total loss function of the model during the finetuning stage consists of two terms:

$$
\begin{aligned}
L_{\text{total}} &= L_{\text{lab}} + \lambda L_{\text{ent}} \\
&= - \mathop{\mathbb{E}}_{\mathbf{S} \in S_l} [log\ p_{\text{model}}(y|\mathbf{S})] \\
&\quad - \lambda \mathop{\mathbb{E}}_{\mathbf{S} \in S_{ul}} \left[\sum_{k=1}^{K} p_{\text{model}}(y|\mathbf{S})\ log\ p_{\text{model}}(y|\mathbf{S}) \right]
\end{aligned}
$$

where L_{lab} corresponds to the standard cross-entropy loss and L_{ent} to the entropy minimisation penalty. A hyperparameter λ is used to regulate the contribution of the unlabeled samples to the total loss function. Under this setup, the model is trained for an additional E_2 epochs, using both labeled and unlabeled samples.

C. Intake cycle detection

Given a trained model and a sequence $\tilde{\mathbf{S}}$ representing a new meal session, food intake cycles can be detected by using a sliding window approach. A window of length W_l traverses the score sequence with a shift of W_s and at each step, which is associated with the timestamp of the last sample of the window, feeds the extracted sub-sequence to the trained LSTM. In this way, the probability of food intake cycle is computed for each timestamp, resulting in a 1-dimensional signal of intake probability versus time. The signal is smoothed using a median filter to remove very sharp peaks and subsequently passed through a differentiation filter to remove large peaks that are too close to each other. Finally, food intake cycles are identified as the local maxima of the filtered signal that are above a detection threshold, T_d.

III. Dataset

A. Food Intake Cycle Dataset (FIC)

In this work, we use the publicly available FIC dataset. The FIC dataset contains one meal session recording for 10 different subjects. During each meal, the accelerometer and gyroscope streams were captured with a sampling rate of approximately 62Hz, using a Microsoft Band 2 smartwatch.

The ground truth was created by associating each sensor-generated sample $s(t)$ with the micro-movement the subject was performing at that t. To this end, an auxiliary video recording of the meal session was used. The micro-movement annotations can be used to form the intake cycle ground-truth, by stipulating that a food intake cycle is a sequence of micro-movements that always begins with the micro-movement "pick food", contains the micro-movement "mouth" and ends with the micro-movement "downwards". The collection protocol, as well as, a detailed description of the annotation procedure are provided in [1]. The dataset is publicly available at `https://mug.ee.auth.gr/intake-cycle-detection/`.

B. Extended FIC

In addition to the FIC data, we collected additional recordings to use as unlabeled samples. In particular, we collected 2 additional recordings from 5 of the 10 subjects of the original FIC corpus, thus bringing the total number of recorded sessions for these subjects to three. Due to deprecation, the Microsoft Band 2 was replaced with a Sony smartwatch, but the capturing conditions (venue, types of food, etc.) were otherwise identical to the original recordings. The new sessions were annotated at a food intake cycle level using the video recording. This ground truth was not revealed to our algorithm, but was merely used to segment a session into relevant sub-sequences that is, sub-sequences that belong exclusively to one of the two classes of interest.

This additional data collection process resulted in an enhanced version of the FIC dataset. This new version contains multiple meal sessions for 5 subjects of the original corpus that will be used to accommodate the needs of the semi-supervised approach. The new dataset will soon be made publicly available at the same location as the original FIC.

IV. Experiments & Results

Training and evaluation was performed by employing a *Leave One Subject Out (LOSO)* scheme. Specifically, for each of the 5 subjects in our extended dataset, we trained a baseline model for 30 epochs using the labeled sessions of the rest. The weights of the model after this first training stage were used as the starting point of the finetuning stage. This stage used two of the three available sessions of the left-out subject and ran for 20 epochs. Training of the baseline model also continued for the same number of epochs so that both the baseline and the corresponding finetuned model receive the same amount of training in order to allow their proper comparison. Evaluation was carried out in the session of the subject which was left out during finetuning. This process was repeated 3 times per left-out

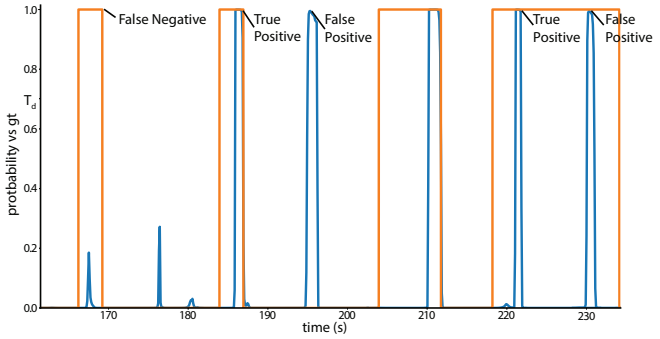

Fig. 2: Evaluation framework overview.

subject in order to examine all the possible training-testing splits of the unlabeled data (that is, 2 sessions for finetuning and 1 session for evaluation). Each of the LOSO iterations described above was repeated 10 times for all of the 5 subjects that contributed their data to the extended version of the FIC dataset. Both the baseline and the finetuned model were trained under the same architecture, including any modifications mentioned in section II.

The unlabeled loss weight, λ, was set to 0.2, so that the contribution of each term in the loss function would be roughly of the same order. The sliding window length, W_l, was set to $7s$ and its shift, W_s, was $0.2s$ at each step. The detection threshold, T_d, was set to 0.7 for all subjects after experimenting with a small subset of the data.

Given the detection methodology of Section II.C, a detection peak was considered true positive if it occured within a positive ground-truth region. However, when more than one peaks occured within the same positive ground-truth region (orange rectangle in Fig. 2), we consider only the first as a true positive and the rest as false positives (see Fig. 2).

According to this evaluation framework, the performance metrics of our approach for each subject of the LOSO scheme were computed. A comparison with the baseline approach is provided in TABLE II.

	Baseline			**Finetuning**		
Subject	Precision	Recall	F-score	Precision	Recall	F-score
1	0.913	0.750	0.822	0.890	0.831	0.860
2	0.955	0.986	0.970	0.952	0.986	0.969
3	0.951	0.876	0.911	0.942	0.916	0.929
4	0.899	0.951	0.923	0.882	0.955	0.917
5	0.964	0.881	0.919	0.962	0.892	0.925
Average	0.936	0.888	0.909	0.925	0.916	0.920

TABLE II: Comparison of performance metrics obtained by the baseline and the proposed method, averaged over 10 independent trials.

Based on these early results, we notice that the finetuning process results in increased recall scores accompanied by a small tradeoff in precision, thus leading to an overall increase in F-score in most cases. Moreover, we notice that subjects whose baseline performance is lower (e.g. subject 1), tend to benefit more from the finetuning stage than others for whom the baseline model was already performing well (e.g. subject 2). This trend highlights the potential of obtaining meaningful improvements in performance for subjects whose eating style is underepresented in the original training set and for which, consequently, a pre-trained model will likely fall short. Finally, it is worth mentioning that if the food type of the evaluation meal was represented in the data used for finetuning then the latter was more likely to succeed.

V. Conclusions

We have presented a method for adapting a pre-trained IMU-based food intake cycle detection model to a new subject using a semi-supervised learning approach. Early results indicate that unlabeled samples can indeed be used to correct the decision boundaries of the pre-trained model in unexplored regions of space. Evaluation on an extension of the FIC dataset, shows improvements of the proposed method over the baseline on average. Collection of a significantly larger dataset with multiple meals per subject is underway in order to fully explore the potential of the method.

VI. Acknowledgments

The work leading to these results has received funding from the EU Commission under Grant Agreement No. 727688 (http://bigoprogram.eu, H2020). We also thank the NVIDIA Corporation for the donation of the Tesla K40 GPU that was used in this research.

References

[1] K. Kyritsis, C. L. Tatli, C. Diou, and A. Delopoulos, "Automated analysis of in meal eating behavior using a commercial wristband imu sensor," in *Proceedings of 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 07 2017, pp. 2843–2846.

[2] M. Zandian *et al.*, "Linear eaters turned decelerated: Reduction of a risk for disordered eating?" *Physiology and Behavior*, vol. 96, no. 4, pp. 518 – 521, 2009.

[3] C. Maramis, C. Diou, I. Ioakeimidis, I. Lekka, G. Dudnik, M. Mars, N. Maglaveras, C. Bergh, and A. Delopoulos, "Preventing obesity and eating disorders through behavioural modifications: The SPLENDID vision," in *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, Nov 2014, pp. 7–10.

[4] V. Papapanagiotou, C. Diou, B. Langlet, I. Ioakimidis, and A. Delopoulos, "Automated extraction of food intake indicators from continuous meal weight measurements," in *Bioinformatics and Biomedical Engineering*, Cham, 2015, pp. 35–46.

[5] V. Papapanagiotou, C. Diou, L. Zhou, J. van den Boer, M. Mars, and A. Delopoulos, "A novel chewing detection system based on ppg, audio, and accelerometry," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 607–618, May 2017.

[6] Y. Dong *et al.*, "A new method for measuring meal intake in humans via automated wrist motion tracking," *Applied psychophysiology and biofeedback*, vol. 37, no. 3, pp. 205–215, 2012.

[7] K. Kyritsis, C. Diou, and A. Delopoulos, "Food intake detection from inertial sensors using lstm networks," in *New Trends in Image Analysis and Processing – ICIAP 2017: ICIAP International Workshops*, Catania, Italy, 09 2017, pp. 411–418.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computing*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[9] I. Karakostas, V. Papapanagiotou, and A. Delopoulos, "Building parsimonious svm models for chewing detection and adapting them to the user," in *New Trends in Image Analysis and Processing – ICIAP 2017*, Catania, Italy, 2017, pp. 403–410.

[10] O. Chapelle, B. Schlkopf, and A. Zien, *Semi-Supervised Learning*, 1st ed. The MIT Press, 2010.

[11] Y. Grandvalet *et al.*, "Semi-supervised learning by entropy minimization," in *Proc of the 17th Intl Conf on Neural Information Processing Systems*, ser. NIPS'04, 2004, pp. 529–536.