PAPER

Smart Tableware-Based Meal Information Recognition by Comparing Supervised Learning and Multi-Instance Learning

Liyang ZHANG^{†a)}, Hiroyuki SUZUKI[†], *Nonmembers*, and Akio KOYAMA[†], *Member*

SUMMARY In recent years, with the improvement of health awareness, people have paid more and more attention to proper meal. Existing research has shown that a proper meal can help people prevent lifestyle diseases such as diabetes. In this research, by attaching sensors to the tableware, the information during the meal can be captured, and after processing and analyzing it, the meal information, such as time and sequence of meal, can be obtained. This paper introduces how to use supervised learning and multi-instance learning to deal with meal information and a detailed comparison is made. Three supervised learning algorithms and two multi-instance learning algorithms are used in the experiment. The experimental results showed that although the supervised learning algorithms have achieved good results in F-score, the multi-instance learning algorithms have achieved better results not only in accuracy but also in F-score.

key words: meal information, smart tableware, Internet of Things, supervised learning, multi-instance learning

1. Introduction

Lifestyle diseases are defined as diseases related to the way a person or group of people lives, including atherosclerosis, heart disease, obesity and type 2 diabetes, and are associated with alcohol, drugs and smoking abuse, as well as lack of physical activity and unhealthy diet [1]. About 13% of the adult population worldwide was obese in 2016 [2]. The number of people with diabetes was 108 million in 1980, increased to 422 million in 2014 [3]. Simple lifestyle measures have been shown to effectively prevent or delay the onset of these diseases, for example, healthy weight maintenance, regular exercise, healthy meal and so on.

In recent years, people pay more and more attention to how to live a healthy life, and with the popularity of Internet of Things, devices that monitor food intake have attracted our attention. Some methods monitor food intake from chewing, swallowing, and gestures by using wearable sensors such as piezoelectric sensors, gyroscopes, and accelerometers. To estimate energy intake, a method of detecting the chewing episodes by a piezoelectric film sensor in a meal was presented [4]. The sensor is attached below the outer ear to detect changes in skin curvature during chewing. In the evaluation of this research the F-score for food intake detection was about 91.0%. However, the drawback of this research is that attaching the sensor under the outer ear may make users feel uncomfortable. Food intake monitoring sys-

tem based on wearable wireless necklace was proposed [5], which can capture the motion in the throat through the embedded piezoelectric sensor in the necklace. The result of F-score was about 80%, but the drawback of this research is that the tightness of the necklace will affect the monitoring results. To find the correlation between eating activities and wrist motion, Dong et al. [6] tracked the linear and rotational motion of wrist by wearing watch-like configuration of accelerometers and gyroscopes. In this research, the accuracy of detecting eating reached 81%.

At the same time, the use of cameras to record people's lives has become popular. FoodLog [7] is a photo-based multimedia system that records food intake by taking pictures of meals. There are also systems [8] that automatically segment and recognize different foods on a plate by taking food images on a smartphone. A method of measuring food intake by using Kinect camera was proposed [9]. It can measure user's food intake by detecting the hand wrist rotation and movement of jaw during meal. There may be overlapping of hand and face detection, and the overall accuracy was about 94%.

In addition, compared with focusing on the amount and composition of the meal, studies found that for type 2 diabetes, eating vegetables, whey protein and olive oil before carbohydrate can reduce increasing of glycemic [10], [11]. A proper sequence of meal can help prevent some of the lifestyle diseases, such as diabetes. Therefore, we need to pay attention to a proper sequence of the meal.

In our previous studies, we used an acceleration sensor attached to tableware to collect meal information [12]. However, more detailed information cannot be obtained only through the acceleration sensor. So, in order to obtain more detailed information, we used multiple sensors like a weight sensor and an acceleration sensor attached to the tableware and applied different algorithms [13], [14]. Through experiments, we found that although smart tableware can sense the weight change and the movement state of the tableware, it cannot sense the chewing process, which makes it difficult to judge whether it is chewing when using the supervised learning algorithms, and thus it is difficult to judge the meal time accurately. Therefore, in this research, we consider using multi-instance learning (MIL) [15] to organize the instances of multiple actions during the meal into bags and classify the bags. The experimental results show that, compared with supervised learning, MIL can achieve higher accuracy and F-score.

This paper will briefly introduce the smart tableware,

Manuscript received March 3, 2020.

Manuscript revised August 17, 2020.

Manuscript publicized September 18, 2020.

[†]The authors are with Graduate School of Science and Engineering, Yamagata University, Yonezawa-shi, 992–8510 Japan.

a) E-mail: ttf04338@st.yamagata-u.ac.jp

DOI: 10.1587/transinf.2020EDP7043

which senses the process of meal through sensors attached to it. At the same time, we will introduce in detail how to use the MIL and supervised learning to predict the meal information in this research and compare the MIL algorithms with the supervised learning algorithms through experiments.

The rest of paper is organized as follows. In Sect. 2, we will introduce the structure of smart tableware, feature extraction, and how to use MIL and supervised learning. The proposed methodology will be evaluated in Sect. 3. Finally, the conclusion of this paper is given in Sect. 4.

2. Methodology

In order to get meal information, we attach multiple sensors to the tableware. Each sensor has an identification number. By assigning food to the sensor attached the tableware in advance, we can determine what food is eaten according to the identification number of smart tableware. During the meal, we use the information of the smart tableware captured by the sensors, and use machine learning to determine whether the food is eaten. We have also developed the meal information collection system [13] for users to browse. The data obtained from the smart tableware will be transmitted to a local PC via wireless LAN, then the local PC transmits the data to the server attached to the Internet. Users can browse the meal information, and the system will give some suggestions based on the analysis of these information to help people prevent lifestyle diseases. In this paper, we focus on the comparison between the MIL algorithms and the supervised learning algorithms.

2.1 Smart Tableware

For smart tableware, we currently use two sensors, namely an acceleration sensor and a pressure sensor. We use the “IoT wireless tag TWE-Lite-2525A” manufactured by Mono Wireless Inc. to capture the acceleration information, and the pressure information is captured using “pressure sensor AS-FS” manufactured by AsakuasGiken Co., LTD. Figure 1 shows the location of the sensors attached to the smart tableware. We attach the acceleration sensor to the tableware and place the tableware on the pressure sensor for

Fig. 1 Location of sensors attached to tableware.

experiments. According to the configuration of the smart tableware and the information that the tableware can capture, we now focus on three types of information: ID, acceleration information and pressure information.

1) ID

Each sensor has an ID, and sensors with different IDs will be attached to the tableware, so that we can distinguish food information by ID. Here we will map the food to the tableware in advance, so that different foods will be mapped to different IDs through the tableware. In this way, not only the name of the food, the corresponding calories, nutrients and other information can also be known through the ID.

2) Acceleration information

The movement of the tableware can be described by acceleration information, which is captured by the acceleration sensor. Here we give the information form that can be captured by the acceleration sensor, as shown in Fig. 2, which contains the current time, ID of the sensor, action mode, etc. The action mode “0008” indicates that the sensor senses that the tableware is moving. An action mode of “0010” indicates that the tableware is stationary.

3) Pressure information

We use the pressure sensor to capture changes in the weight of the food in the tableware. The pressure sensor we use is a voltage change type sensor, which converts the load (the weight of the food) into voltage value, and the output voltage increases when the load increases. The information form that can be captured by the pressure sensor is shown in Fig. 3, which contains the current time, ID of the sensor, voltage value, etc.

We want to judge whether the food is being eaten or not by judging whether the tableware is moving and whether the weight of food is decreasing. At the same time, combining the information of ID, we can confirm the sequence and time of meal, as shown in Fig. 4.

2.2 Feature Extraction

It is not easy if we use the information captured from the sensors directly, which may not reflect the continuous

Time ID Action Mode
;210;000000000;177;075;1020e89;2990;0008;0000;1302;0651;X;0067;-053;0021;

Fig. 2 An example of information format obtained from acceleration sensor.

Time ID Voltage Value
;99;000000000;135;262;10f1780;3070;0808;7392;1232;0808;S;

Fig. 3 An example of information format obtained from pressure sensor.

Fig. 4 Information captured by smart tableware.

Fig. 5 Feature extraction based on time sequence.

Fig. 6 An example of process of meal.

change process of state of tableware. So far we have extracted five features, including two acceleration features and three pressure features. The two acceleration features extracted are “the detection count above the threshold” and “the continuous number above the threshold”. The three pressure features extracted are “pressure difference (voltage difference)”, “continuous number above pressure difference average” and “deviation rate of moving average”. We suppose that these extracted acceleration features can reflect whether the tableware is in the state of motion, and whether it is in the state of continuous motion, and these extracted pressure features can reflect the process and trend of food weight change. In the experiment, we extract the feature vector every 5 s according to the time sequence, as shown in Fig. 5.

2.3 Multi-Instance Learning (MIL)

Through experiments, we suppose that the following steps are generally included in the process of meal with the smart tableware (Fig. 6). Step 1: start to pick up food from tableware with chopsticks; Step 2: send food to the mouth; Step 3: chew.

We know that in step 1 and step 2, the weight of the food in the tableware will be changed or there will be a process of moving the tableware to the mouth or away from the mouth, which can be sensed by sensors, but step 3 cannot be well sensed by sensors attached to tableware. It is not easy to judge the data of step 3 accurately when using supervised learning algorithms to predict “eat” and “not eat”. At the same time, we can capture the time from the start to the end of the meal, but cannot judge the actual eating time accurately. Therefore, we consider using MIL to deal with this problem.

MIL considers a special form of weak supervision, and it has been applied to many fields. In MIL, the training set consists of a set of positive bags and a set of negative bags, and each bag contains one or more instances. The label corresponds to the bag. A bag is classified as positive if it contains at least one instance related to the subject of interest,

Fig. 7 An example of meal sequence and time based on MIL.

and a bag is classified as negative if it does not contain any instances related to the subject of interest. MIL is to train a classifier on the labeled bags and use the classifier to predict unknown bags.

Through the data obtained from the smart tableware, we need to distinguish “eat” or “not eat” according to the time sequence. Based on the description of MIL, instances related to “eat”, that is, from taking food with chopsticks or a spoon to chewing, are considered as instances in the positive bag. And instances related to “not eat” are regarded as instances in the negative bag. Figure 7 shows an example of a 40-second meal process. According to the time sequence, a feature vector is extracted every 5 seconds, and a total of eight feature vectors are extracted. We regard the process of drinking a bite of miso soup as a positive bag, and the process of eating a bite of rice as another positive bag. There is a break between drinking miso soup and eating rice. Therefore, according to this process, the meal sequence can be judged as drinking miso soup first and then eating rice, the time to drink miso soup is 15 s, and the time to eat rice is 20 s. In this way, meal information such as meal sequence and meal time can be obtained.

An object is represented by an instance (i.e., a feature vector) and associated with a label in supervised learning, but in MIL an object is described by multiple instances and associated with a label [16]. There are many ways to solve MIL. In this part, we try to map each bag to a feature vector to summarize the relevant information in the bag, then standard supervised learning can be used to solve this classification problem. A bag can be represented by the sum of all instances in it, normalized by its 1 or 2-norm, in Normalized Set Kernel (NSK) [15] algorithm, and then trained by Support Vector Machine (SVM). Suppose there are n instances in a bag, where v_i ($1 \leq i \leq n$) represents the feature vector of the i -th instance in it. For the NSK algorithm, we can understand it in this way: first, by summing up all instances in the bag, $v = v_1 + \dots + v_i + \dots + v_n$ can be obtained. Here, suppose that v is a five-dimensional vector $(x_1, x_2, x_3, x_4, x_5)$. Then normalized by its 1 or 2-norm $(x_1^p + x_2^p + x_3^p + x_4^p + x_5^p)^{1/p}$ ($p = 1, \text{ or } 2$). After all the bags in the data set are processed in this way, SVM is used to train them. By extracting information from the whole bag, the bag can be represented more comprehensively. Therefore, NSK is used for experiment in this paper. In the experimental results, NSK1 is used to represent 1-norm and NSK2 is used to represent 2-norm.

2.4 Supervised Learning

We will introduce how to use traditional supervised learning

Fig. 8 An example of label setting based on supervised learning.

Fig. 9 An example of meal sequence and time based on supervised learning.

to process meal information in this section. In traditional supervised learning, an instance corresponds to a label, while in MIL, a bag corresponds to a label. As we introduced, in the process of meal, the following 3 steps are generally included: step 1 is to start taking food with chopsticks; step 2 is to use chopsticks to put food into the mouth; step 3 is chewing. We know that the food is indeed eaten in the step 3, while in step 1 and step 2, the food is just about to be eaten. However, to compare with MIL, we set label as “eat” if an instance contains information of any of these 3 steps, and an instance that does not contain information of these 3 steps corresponds to “not eat”, as shown in Fig. 8.

Figure 9 shows an example of process of obtaining meal sequence and time based on supervised learning. We can know that the sequence of meal is to drink the miso soup first and then eat rice, the time of drinking miso soup is 15 s, and the time of eating rice is 20 s. There do not have the concept of bag in Fig. 9, and each instance corresponds to a label. Here, we choose three classical supervised learning algorithms: decision tree, SVM and Bayes network for experimental comparison.

3. Evaluation

In the experiment, we use smart tableware with multiple sensors for the meal experiments and assume that food corresponds to smart tableware. At the same time, we used a video camera to record the meal process to facilitate labeling. In this paper, three supervised learning algorithms and two MIL algorithms are selected, and WEKA [17] is used. 10-fold cross-validation is selected as evaluation approach in the experiment, and the experimental results are compared.

To facilitate the evaluation, a total of 10 meal experiments were performed, and 480 bags and 1205 instances were obtained, as shown in Table 1. We can see that the average number of instances in the positive bag is 5.3, the average number of instances in the negative bag is 1, and the average number of instances in the bag is 2.5.

Table 1 Details of the experimental data set.

Attribute	Value
Number of Bags	480
Number of Instances	1205
Average number of instances in bag	2.5
Average number of instances in positive bag	5.3
Average number of instances in negative bag	1

Table 2 The precision, recall, F-score and accuracy of supervised learning.

	Precision	Recall	F-score	Accuracy
SMO	76.6%	93.7%	84.3%	74.0%
J48	86.7%	83.2%	84.9%	78.1%
BayesNet	86.2%	85.9%	86.1%	79.3%

Here, we use Precision, Recall, F-score and Accuracy to evaluate performance, as shown in Eqs. (1), (2), (3) and (4). In supervised learning, Precision is used to describe the ratio between the number of instances correctly classified as “eat” and the number of instances predicted as “eat”, Recall indicates the ratio between the number of instances correctly classified as “eat” and the number of instances actually “eat”, F-score is the harmonic mean of precision and recall, and Accuracy represents the proportion of correctly classified instances to all instances. Supervised learning is to train the classifier on the labeled instances and use the classifier to predict the unknown instances. In MIL, we evaluate based on bags, T_p denotes the number of bags correctly classified as “eat”, $T_p + F_p$ is the number of bags predicted to be “eat”, $T_p + F_n$ represents the number of bags actually “eat”, and accuracy is the ratio between correctly classified bags and all bags. MIL is to train a classifier on the labeled bags and use the classifier to predict the unknown bags in the paper.

$$\text{Precision} = \frac{T_p}{T_p + F_p} \quad (1)$$

$$\text{Recall} = \frac{T_p}{T_p + F_n} \quad (2)$$

$$\text{F-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \quad (4)$$

For supervised learning algorithms, J48, SMO and BayesNet in WEKA are selected. The C4.5 algorithm is implemented in Weka as a classifier called J48, which is an algorithm producing decision tree based on information theory. SMO (Sequential Minimal Optimization) is widely used to solve the quadratic programming problem that occurs during training of SVM. Bayes network is a type of probabilistic graphical model and have become popular in artificial intelligence. The Precision, Recall, F-score and Accuracy of the supervised learning algorithms are shown in Table 2. We can see that BayesNet has the highest F-score of 86.1%, and its Precision and Recall are around 86%. The F-score of SMO and J48 are both about 84%. SMO has a high Recall of 93.7%, but about 23.4% of the instances are incorrectly predicted as “eat”, among the instances pre-

Table 3 The precision, recall, F-score and accuracy of MIL.

	Precision	Recall	F-score	Accuracy
NSK1	89.0%	90.0%	89.5%	92.5%
NSK2	89.8%	88.2%	89.0%	92.3%

dicted as “eat”. We can see that for the Accuracy of supervised learning algorithm, the SMO is the lowest at 74%, the Accuracy of J48 is 78.1%, and BayesNet is 79.3%, which means that the percentage of misclassified instances reaches more than 20%. Among these misclassified instances, some of them should be instances related to chewing, which is related to the fact that smart tableware can capture the state of the tableware but may not be able to judge the chewing accurately.

For MIL algorithms, NSK is used. In NSK, a bag is represented by the sum of all its instances, normalized with its 1 or 2-norm, and then trained with SVM. In the experimental results, NSK1 is used to represent 1-norm and NSK2 is used to represent 2-norm. Table 3 shows the Precision, Recall, F-score and Accuracy of the MIL algorithms. We can see that the Precision of NSK1 and NSK2 are 89.0% and 89.8%, respectively, which means that 11.0% and 10.2% of the bags predicted to be “eat” in NSK1 and NSK2 are misclassified. The Recall of NSK1 and NSK2 are 90.0% and 88.2%, 10.0% and 11.8% of the bags that actually “eat” are predicted as “not eat” in NSK1 and NSK2, respectively. The F-score of NSK1 and NSK2 are 89.5% and 89.0% respectively, which are inseparable from the higher Precision and Recall. Compared with supervised learning, the Accuracy of MIL algorithms NSK1 and NSK2 has been significantly improved, reaching 92.5% and 92.3% respectively, which are both higher than 90%, and the proportion of misclassified bags is less than 10%. In MIL, the bag is represented by the combination of all instances from eating with chopsticks to chewing, and there is no need to predict the chewing periods separately. However, in supervised learning, each instance corresponds to a label, and the chewing part needs to be predicted separately. This may be the reason why MIL has a higher classification accuracy than supervised learning. Based on the analysis of experimental results, MIL is superior to supervised learning in terms of Accuracy and F-score in general, and MIL is better at predicting the time and sequence of meals.

4. Conclusion

This paper introduces the smart tableware and the process of collecting meal information briefly and introduces in detail how to deal with meal information by MIL and supervised learning. In the experiment, we use five algorithms, including three supervised learning algorithms and two MIL algorithms, and compare the Accuracy and F-score of these five algorithms. Compared with the experimental results, although the supervised learning algorithms achieve good results on F-score, the MIL algorithms are superior to supervised learning in both Accuracy and F-score.

In the future, we focus on using sensitive sensors to

capture the state of the smart tableware accurately and extracting more features to reflect the state of the smart tableware. In the future, we hope to make efforts to apply smart tableware so that people can use it in real life.

References

- [1] MedicineNet.com, Medical Definition of lifestyle disease, <https://www.medicinenet.com/script/main/art.asp?articlekey=38316>, accessed Aug. 17. 2020.
- [2] World Health Organization, Obesity and overweight, <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>, accessed Aug. 17. 2020.
- [3] World Health Organization, Diabetes, <https://www.who.int/news-room/fact-sheets/detail/diabetes>, accessed Aug. 17. 2020.
- [4] M. Farooq and E. Sazonov, “Automatic measurement of chew count and chewing rate during food intake,” *Electronics (Basel)*, vol.5, no.4, 62, 2016.
- [5] H. Kalantarian, N. Alshurafa, T. Le, and M. Sarrafzadeh, “Monitoring eating habits using a piezoelectric sensor-based necklace,” *Computers in Biology and Medicine*, vol.58, pp.46–55, 2015.
- [6] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover, “Detecting periods of eating during free-living by tracking wrist motion,” *IEEE J. Biomed. Health Inform.*, vol.18, no.4, pp.1253–1260, 2014.
- [7] K. Aizawa and M. Ogawa, “FoodLog: Multimedia tool for healthcare applications,” *IEEE Multimedia*, vol.22, no.2, pp.4–8, 2015.
- [8] M. Anthimopoulos, J. Dehais, S. Shevchik, B.H. Ransford, D. Duke, P. Diem, and S. Mougiakakou, “Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smart-phones,” *Journal of Diabetes Science and Technology*, vol.9, no.3, pp.507–515, 2015.
- [9] M.F. bin Kassim and M.N.H. Mohd, “Food intake gesture monitoring system based-on depth sensor,” *Bulletin of Electrical Engineering and Informatics*, vol.8, no.2, pp.470–476, 2019.
- [10] H. Kuwata, M. Iwasaki, S. Shimizu, K. Minami, H. Maeda, S. Seino, K. Nakada, C. Nosaka, K. Murotani, T. Kurose, Y. Seino, and D. Yabe, “Meal sequence and glucose excursion, gastric emptying and incretin secretion in type 2 diabetes: A randomized, controlled crossover, exploratory trial,” *Diabetologia*, vol.59, no.3, pp.453–461, 2016.
- [11] S. Imai, M. Matsuda, G. Hasegawa, et al, “A simple meal plan of ‘eating vegetables before carbohydrate’ was more effective for achieving glycemic control than an exchange-based meal plan in Japanese patients with type 2 diabetes,” *Asia Pac. J. Clin. Nutr.*, vol.20, no.2 pp.161–168, 2011.
- [12] K. Kaiya and A. Koyama, “Design and implementation of meal information collection system using IoT wireless tags,” *Proc. 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, Fukuoka, pp.503–508, 2016.
- [13] L. Zhang, K. Kaiya, H. Suzuki, and A. Koyama, “A smart tableware-based meal information collection system using machine learning,” *International Journal of Web and Grid Services*, vol.15, no.2, pp.206–218, 2019.
- [14] L. Zhang, K. Kaiya, H. Suzuki, and A. Koyama, “Meal information recognition based on smart tableware using multiple instance learning,” *Proc. 22nd International Conference on Network-Based Information Systems, NBIS-2019, AISC 1036*, pp.189–199, 2019.
- [15] R.C. Bunescu and R.J. Mooney, “Multiple instance learning for sparse positive bags,” *Proc. 24th International Conference on Machine Learning*, Corvallis, OR, pp.105–112, 2007.
- [16] Z.-H. Zhou, M.-L. Zhang, S.-J. Huang, and Y.-F. Li, “MIML: A framework for learning with ambiguous objects,” *CoRR* abs/0808.3231, 2008.
- [17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, “The WEKA data mining software: An update,” *SIGKDD Explorations Newsletter*, vol.11, no.1, pp.10–18, 2009.

Liyang Zhang received her B.E. degree from Xi'an University of Technology in 2013 and M.E. degree from Guangdong University of Technology in 2016. From April 2018, she is a Ph.D. student at the Graduate School of Science and Engineering, Yamagata University, Japan. Her research interests include Internet of Things and machine learning algorithms.

Hiroyuki Suzuki received his B.E. and M.E. degrees from Yamagata University in 2010 and 2012, respectively. From October 2017, he is a Ph.D. student at Graduate School of Science and Engineering, Yamagata University, Japan. His research interests include application development method of IoT and network systems.

Akio Koyama is a Professor at Department of Informatics, Yamagata University, Japan. He received B.E. and Ph.D. degrees from Yamagata University in 1987 and 1998, respectively. He has published over 100 papers in Journals, Books and International Conference Proceedings. He has been a PC Member of many international conferences and has served as General Co-Chair of NBIS-2012 and Workshops Co-Chair of AINA-2008, CISIS-2009, NBIS-2010 and BWCCA-2010. His research interests include

routing protocols, ad-hoc and sensor networks, agent-based systems, medical and health care systems and IoT/M2M systems. He is a senior member of IEEE, and members of IEICE and IPSJ.