Micro-environment Recognition in the context of Environmental Crowdsensing

Mohammad Abboud
DAVID Lab
UVSQ - Université Paris-Saclay
Versailles, France
mohammad.abboud.2496@gmail.com

Hafsa El Hafyani
DAVID Lab
UVSQ - Université Paris-Saclay
Versailles, France
hafsa.el-hafyani@uvsq.fr

Jingwei Zuo
DAVID Lab
UVSQ - Université Paris-Saclay
Versailles, France
jingwei.zuo@uvsq.fr

Karine Zeitouni
DAVID Lab
UVSQ - Université Paris-Saclay
Versailles, France
karine.zeitouni@uvsq.fr

Yehia Taher
DAVID Lab
UVSQ - Université Paris-Saclay
Versailles, France
yehia.taher@uvsq.fr

ABSTRACT

With the rapid advancements of sensor technologies and mobile computing, Mobile Crowd-Sensing (MCS) has emerged as a new paradigm to collect massive-scale rich trajectory data. Nomadic sensors empower people and objects with the capability of reporting and sharing observations on their state, their behavior and/or their surrounding environments. Processing and mining multi-source sensor data in MCS raise several challenges due to their multi-dimensional nature where the measured parameters (i.e., dimensions) may differ in terms of quality, variabilty, and time scale. We consider the context of air quality MCS, and focus on the task of mining the context from the MCS data. Relating the measures to their context is crucial to interpret them and analyse the participant's exposure. This paper investigates the feasibility of recognizing the human's context (called herein micro-environment) in an environmental MCS scenario. We put forward a multi-view learning approach, that we adapt to our context, and implement it along with other time series classification approaches. The experimental results, applied to real MCS data, not only confirm the power of MCS data in characterizing the micro-environment, but also show a moderate impact of the integration of mobility data in this recognition. Furthermore, multi-view learning shows similar performance as the reference deep learning algorithm, without requiring specific hardware.

KEYWORDS

Activity Recognition, Multivariate Time Series Classification, Multi-view Learning, Mobile Crowd Sensing, Air Quality Monitoring

1 INTRODUCTION

Nowadays, the Internet of Things (IoT) basically relies on advanced sensor technologies to bridge the physical world and information systems. In particular, along with the widespread use of GPS, various mobile sensors bring rich information collected from both the surrounding environment and human activities, which are generally represented as Geo-referenced Time Series (GTS). Mobile Crowed Sensing (MCS) [12] emerges as a new paradigm, which empowers volunteers to contribute data (i.e., GTS) acquired by their personal sensor-enhanced mobile devices. Polluscope[1], a French project deployed in Île-de-France (i.e., Paris region), is a typical use case study on MCS. It aims at getting insights constantly on individual exposure to pollution everywhere (indoor and outdoor), while enriching the traditional monitoring system with the collected data by the crowd. The recruited participants, on a voluntary basis, collect air quality measurements. Each participant is equipped with a sensor kit, and a mobile device which allows the transmission of collected measurements together with the GPS coordinates as a geo-referenced data stream containing (timestamp, longitude, and latitude). In addition, the participants are asked to annotate their environment type through a custom mobile application. This will allow participants to have personalized insights about their exposure to pollution everywhere, either in indoor and outdoor environments (e.g., Home, Work, Transportation, Streets, Park, etc.), and at a higher resolution along their trajectories, thereby, allowing to capture local variability and peaks of pollution, depending on participants' whereabouts, i.e., *micro-environments*.

It is worth mentioning that air quality strongly depends on the context[2], and so is the individual exposure to pollution. For this reason, there is a great interest of making exposure analysis context-aware. However, the context annotation is by far the most difficult information to collect in a real-life application setting, since a very few participants thoroughly annotate their micro-environment. Therefore, there is a great interest in unburdening the participants by automatically detecting the context.

When exploring visually the data, we noticed that micro-environments preserve a certain pattern. Besides, we observe the existence of an inter-sensor correlation and with the context. Figure 1 shows the evolution of three dimensions (i.e. Black Carbon (BC), NO2 and Particulate Matters (PM)) with micro-environments identification. As shown in Figure 1, BC and NO2 preserve the same shapes and statistical characteristics in the micro-environment "Car". Plus, we note that PM values keep the same statistical characteristics in the micro-environment "Indoor". Moreover, we can observe the existence of a correlation between the three dimensions during the whole timeline.

The idea we promote in this paper is to utilize a wisely chosen annotated dataset, in order to train a model on all the combination

© 2021 Copyright for this paper by its author(s). Published in the Workshop Proceedings of the EDBT/ICDT 2021 Joint Conference (March 23–26, 2021, Nicosia, Cyprus) on CEUR-WS.org. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

[1] http://polluscope.uvsq.fr
[2] In this paper, the terms "context" and "micro-environment" are used interchangeably.

of air quality, and mobility dimensions as predictors of the micro-environment. We hypothesize that the multi-variate time series collected by the MCS campaigns not only depends on the micro-environment but could be a proxy of it.

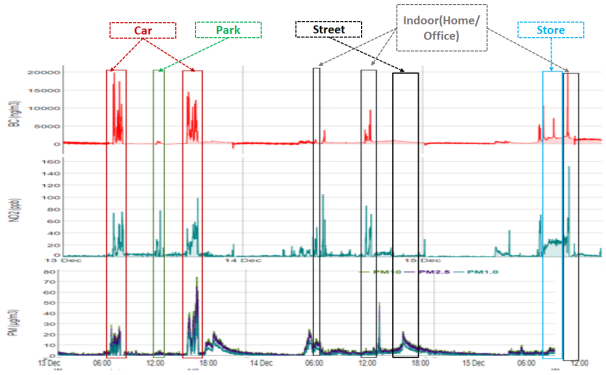


Figure 1: Inter-sensor and micro-environment correlations.

The question that arises itself now is how to combine all these different aspects of the data (geo-location, sensors) to identify the user's context automatically? And how much a model can discriminate the observations in different micro-environments? To this end, we envision a holistic approach of activity recognition, as depicted in [7].

The micro-environment recognition is a crucial figure to exposure interpretation. Once the data are correctly annotated, our ultimate goal is to get insight into all the dimensions (spatial, temporal, individual, and contextual) of the exposure to pollution.

In this paper, we evaluate different approaches and provide a framework dedicated to the preparation, the application and the comparison of different machine learning algorithms.

The rest of this paper is organized as follows. We introduce the related work in Section 2. The formal presentation of our micro-environment recognition model is discussed in Section 3. Section 4 presents the experimental results and evaluation of the micro-environment recognition model in the context of environmental crowd sensing. Section 5 gives the extensive discussion for the perspectives of this work. In section 6, we summarize our conclusions and provide directions for future work.

2 RELATED WORK

Human activity recognition involves a wide range of applications from smart homes activities [1] to daily human activities [4, 17, 28], to human mobility [6, 30] to cite a few. It represents a typical scenario of machine learning, and some public datasets are widely used in the benchmarks. In this section, we introduce a summary of two main topics of related work to our approach. We focus mainly on multi-variate time series (MTS) classification and multi-view learning.

2.1 Multi-Variate Time Series Classification

Human activity recognition falls in the problem of labelling data segments with the type of activity, which leads to a multi-variate time series classification (MTSC) problem based on data collected by multiple wearable sensors. There is a wide range of time series classification approaches that can be classified into four categories: distance-based methods [2], feature-based methods [19], ensemble methods [11] and deep learning models [3, 9, 24]. The One-Nearest Neighbor (1-NN) classifier with different distance measures, such as Euclidean Distance (ED) or Dynamic Time Wrapping (DTW) [2], is always considered as the benchmark to give a preliminary evaluation in the MTSC problem.

Considering the real-life scenarios, where it is difficult or expensive to obtain a large amount of labeled data for training, some studies used both labeled and unlabeled data to learn the human activity, that is Semi-Supervised Learning (SSL) [25] on MTSC. The pioneering work by [25] propose a semi-supervised technique for time series classification. The authors demonstrated that semi-supervised learning requires less human effort and generally achieves higher accuracy than training on limited labels. The semi-supervised model [25] is based on the Self-Learning concept with the One-Nearest-Neighbor (1-NN) classifier. First, the labeled set, denoted by P (as positively labeled) is applied to train the 1-NN classifier C. Then, the unlabeled samples U are given the pseudo labels progressively based on their distance to the samples in P. Thereafter, the enriched labeled set P allows iteratively repeating the previous step and improving the classifier. More recently, the deep learning-based models on MTSC show promising performance under weak supervision. For instance, Zhang et al. [29] propose a novel semi-supervised MTSC model named Time series attentional prototype network (TapNet), to explore the valuable information in the unlabeled samples. TapNet projects the raw MTS data into a low-dimensional representation space. The unlabeled samples approach themselves to the class prototype in the representation space, where the distance-based probability and the labeled samples allow training the model progressively. Moreover, the hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) structure adopted in TapNet allows modeling, respectively, the variable interactions and the temporal features of MTS.

2.2 Multi-View Learning

Another line of studies propose multi-view learning to classify time series data originated from multiple sensors to recognize users activities. Garcia-Ceja et al. [11] propose a method based on multi-view learning and stacked generalization for fusing audio and accelerometer sensor data for human activity recognition using wearable devices. Each sensor's data is seen as a different "view", and they are combined using stacked generalization [26]. The approach trains a specific classification model over each view and an extra meta-learner using the view models as input. The general idea of the authors is to combine data from heterogeneous types of sensors to complement each other and thus, increase recognition accuracy.

Wang et al. [23] propose a framework based on deep learning to learn features from different aspects of the data based on features of sequence and visualization. In order to imitate the human brain, which can classify data based on visualization, the authors transform the time series into an Area Graph. They use well-trained LSTM-A neural networks and CNN-A neural networks to extract the features of time series data. LSTM-A is used to extract sequence features, while CNN-A is used to extract visual features from the time series. Then, based on the fusion of features, the authors carry out the time series classification task. Although the approach gained promising results, it did not outperform deep learning methods such as InceptionTime [10].

Li et al. [15] propose a Multi-view Discriminative Bilinear Projections (MDBP) for multi-view MTSC. The proposed approach is a multi-view dimensionality reduction method for time series

classification which aims to extract discriminative features from multi-view MTS data. MDBP mainly projects multi-view data to a shared subspace through view-specific bilinear projections that preserve the temporal structure of MTS, and learns discriminative features by incorporating a novel supervised regularization.

3 MICRO-ENVIRONMENT RECOGNITION MODEL

In this section, we provide an overview of our proposed framework for micro-environment recognition in the context of MCS. Our proposed approach contains six steps as shown in Figure 2.

3.1 Data Collection

The first step of our micro-environment recognition process is the data collection. During three campaigns, around one hundred participants have been recruited to collect ambient air measurements along with geo-location for one week 24 hours a day, while performing their daily activities. Each participant carries a multi sensor box and a tablet empowered with GPS chipset. The sensors collect time annotated measurements of Particulate Matter (PM1.0, PM10, PM2.5), nitrogen dioxide (NO_2), Black Carbon (BC), Temperature and Relative Humidity, and the tablet records participants' geo-locations and allows them to annotate their context by using a *self-reporting* mobile app. They report every transition to a micro-environment (e.g., Home, Office, Park, Restaurant, etc.), as well as events, which are temporary activities for a brief period (e.g., Start cooking, Open a window, Close a window, Smoking, Turn on a chimney, etc.).

3.2 Data Preparation

The second step consists of pre-processing the data which is two folds. On the one hand, most sensor data are noisy, and require a prepossessing phase to clean them from irrelevant measurements. We have observed this especially in the GPS (due to signal loss), and in air quality data even though the sensor data quality is a permanent preoccupation of the project, by careful evaluation before their selection, and periodic qualification during the campaign [14]. The sensors for climatic parameters do not show such defects. Therefore, a de-noising process is applied to clean the data. On the other hand, the highest quality sample of annotated data is selected as a baseline to validate the process of micro-environment recognition. The idea is to generalize the micro-environment recognition to all participants' data, by using the model derived from a good-quality dataset.

3.3 Multi-View Learning Model

We were interested in the stack generalization approach proposed in [11], but we have adapted it to best fit for solving our problem. We propose to learn the micro-environment of participants from multi-variate time series through a two-stage model based on multi-view learning. Our multi-view classification approach consists of training a first-level learner on each view (i.e. step 3 in Figure 2), and then train a second-level learner or meta-learner (i.e. step 5 in Figure 2) to combine the output of each view and enhance the global accuracy of the classification. We assume that Y_{it} is a dimension of the n-dimenstional time series $Y_t = (Y_{1t}, Y_{2t}, ..., Y_{it}, ..., Y_{nt})$. In our model, each view V_i, where $V = (V_1, V_2, ..., V_i, ..., Vn)$ is the set of views, represents a dimension Y_{it} of the multi-variate time series Y_t. Thus, we have as many views as dimensions.

Table 1: An example of the new generated dataset D'

First-Level Learners						Associated Prediction Probabilities						True Label
l_1	l_2	...	l_i	...	l_n	p_1	p_2	...	p_i	...	p_n	y

In step 3, the first-level learner takes as input the time series data coming from each view. Then, each view will generate its own predicted labels with associated prediction probabilities with the form $[l_i, p_1, p_2, ..., p_j, ..., p_k, y]$, where l_i is the predicted label of the first-level learner i, p_j is the associated prediction probability for each class j of the k possible classes, and y is the true label.

One of the advantages of multi-view learning is its versatility on first and second level learners' choices. One can flexibly substitute classifier choices as wished. We opt for k-nearest-neighbor (kNN) classifier coupled with the Dynamic Time Warping (DTW) distance as first-level learner to be trained on each view of the data. kNN is one of the most popular and traditional TSC approaches. kNN with DTW metric was considered for a long time the state-of-the-art in the time series classification problem [9]. Furthermore, most classification approaches require parameters settings, whereas kNN with DTW is parameter free. kNN classifier has shown to be a strong baseline [9], and there is no single distance measure that significantly outperforms DTW [9]. Hence, recent research has focused on developing ensemble methods that significantly outperforms the NN coupled with DTW [9].

In step 4, we aimed at giving a weight for each learner, thus a new dataset D' is generated by joining the first-level learner predictions and the probability of each prediction, Table 1 shows the feature vector in this dataset, where l_i is the predicted label of the first-level learner i, p_i is the probability of this prediction, and y is the true label.

In step 5, after generating a new dataset D', a second-level classifier, or meta-learner, is trained over D' through ensemble learning [31]. This approach allows to preserve the statistical properties of each view and learn the classes of the MTS instances with a significant improvement in the classification accuracy.

Many ensemble methods [31] have been proposed to further enhance the algorithm's accuracy by combining learners rather than trying to find the best single learner. Due to their versatility and flexibility, ensemble methods attract many researchers and can be applied in different domains, for example, but not limited to, time series classification [11] and time series segmentation [8]. In a previous work [8], we used a multi-view approach for segmenting MCS data, where we employed an unsupervised learning for change detection on each view.

In this work, we conduct our experiments using *Random Forest* classifier since it has shown high performance when it is applied in the human activity recognition domain [11].

4 EXPERIMENTS AND RESULTS

The experiments are carried out on different environments. The multi-view learning model was implemented in Python 3.6 using scikit-learn 0.23.2 and tslearn [21]. The deep-learning models (MLSTM-FCN [13], TapNet [29]) were trained on a single Tesla V100 GPU of 32 Go memory with CUDA 10.2, using respectively Keras 2.2.4 and PyTorch 1.2.0.

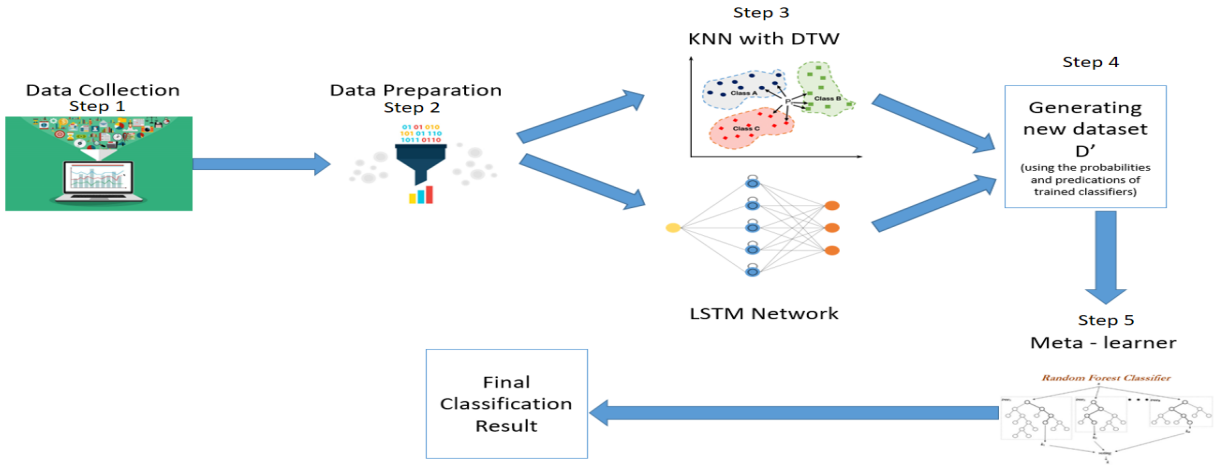


Figure 2: Overview of the Micro-Environment Recognition Process.

4.1 Experimental Settings

We evaluated the used models in these experiments using real life data collected within the scope of Polluscope project. For the experiments we have used the participants' ambient air data containing (PM10, PM1.0, PM2.5, NO2, BC, Temperature, and Relative Humidity), in addition to the speed dimension derived from the geo-locational data. Moreover, we have 8 classes (i.e., micro-environment to recognize), which can be divided into two categories indoor such as ("Home", "Office", "Restaurant", and "Store"), and outdoor such as ("Street", "Bus", "Car", and "Train").

We have selected the data of six participants who have thoroughly annotated their activities within the campaign. Data were split into two third for training and one third for testing, with care taken to keep the data of each participant grouped either in the training or in the testing sets. We used the cross validation score with "Repeated Stratified K-fold" in order to split the training set into training and validation, while we test the overall accuracy using the test dataset.

To account for the temporal feature of the data, we segment them into samples of 10 minutes' length at maximum. Usually, people spend most of their time indoors, thus we should take into consideration other outdoor activities that have a short period of time compared to indoor activities. For example, the average time spent in "Bus" is around 8 minutes, for "Car" the average time is 20 minutes, etc. Globally, the distribution of data samples is highly imbalanced over the different classes, as shown in Figure 3a, which reflects the imbalance of time spent in different micro-environments. Imbalanced classes usually cause low performance classification for the minority classes. To cope with this problem, we apply a resampling strategy. Figure 3 shows the distribution of the data in both the original and the re-sampled dataset. We have used the random over/under-sampler in order to balance our dataset.

To consider the valuable variable of the mobility information, we carry out our experiments on the datasets while considering or not the *speed* variable. We also compare the classifiers' performance on both resampled and original (i.e., un-resampled) data. Finally, we introduce and evaluate a two-steps approach, by first discriminating indoor from outdoor environments, followed by a refinement step to learn a more specific class.

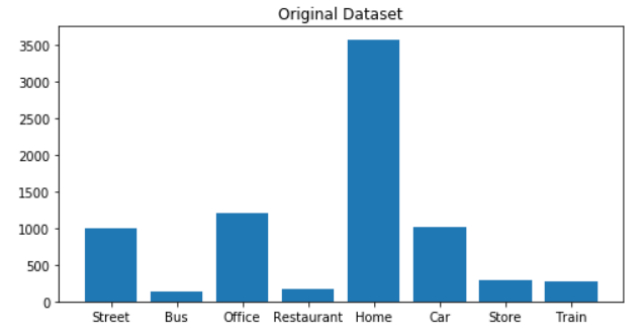
4.2 Classification Results

This section details the experimental results. The micro-environment recognition is formulated as a MTSC problem. We used as a baseline a basic kNN classifier with DTW as distance metric. To compare our multi-view learning approach (*2NN-DTW* for the first-level learners and *Random Forest* as a meta-learner) with state-of-the-art techniques, we implemented the MLSTM-FCN [13] and run it on a GPU, since it requires more computational resources than the multi-view approach. Both kNN-DTW and MLSTM-FCN were applied on an aggregated feature vector containing all the dimensions together.

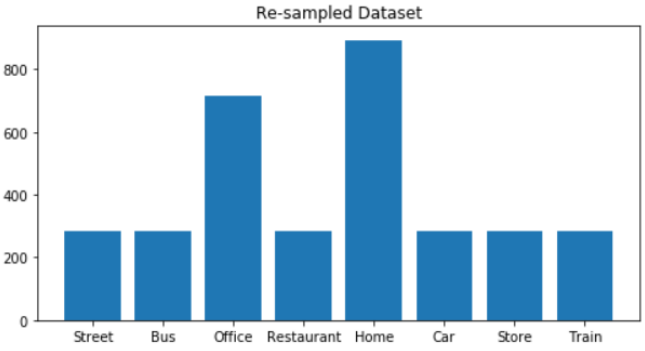
As shown in Table 2, kNN-DTW classifier is not able to discriminate correctly between the micro-environments, while the accuracy improves when applying the multi-view approach that treats each view independently, thus preserving the statistical features of each view. For the third experiment, a Long Short Term Memory network is generated in order to learn a mapping between the input vector and the classes. MLSTM-FCN has shown promising results in the experiments. Table 2 shows the accuracy of experiments carried out with different conditions.

Table 2: Performance of different classifiers on Multi-source Polluscope data

Model	Condition	Accuracy
kNN-DTW	Speed	0.450
	No speed	0.440
	Speed & Re-smp.	0.587
	No speed & Re-smp.	0.597
Multi-view Based	Speed	0.716
	No speed	0.710
	Speed & Re-smp.	0.729
	No speed & Re-smp.	0.640
MLSTM-FCN	Speed	0.808
	No speed	0.784
	Speed & Re-smp.	0.703
	No speed & Re-smp.	0.691
Grouping Step	Speed & Re-smp.	0.83
	No speed & Re-smp.	0.82



(a) Original Dataset



(b) Re-sampled Dataset

Figure 3: Class Distribution

Next, we focus on the comparison between the proposed approach and MLSTM-FCN. We also study the impact of using or not the mobility data, as well as the learning from the original or the re-sampled data. We report the performances in terms of recall, and F1 score. These results are grouped in Table 2 to Table 7 and Figure 4 to Figure 6. Table 2 represents the overall accuracy of the different classifiers used in these experiments while or not using the *Speed* data, and with or without re-sampling. Table 3 and 4 reports the precision, recall, and the F1-Score metrics of the *Multi-view learner* for raw data and re-sampled data respectively with and without Speed. Table 5 and 6 reports the precision, recall, and F1-Score metrics of the *MLSTM-FCN* for raw data and re-sampled data respectively with and without Speed. Figure 4 shows the accuracy among different views used within the experiment of the *Multi-view approach*. Moreover, figure 5a and 5b shows the confusion matrix when applying the *Multi-view approach* on the re-sampled data with/out the speed respectively. Figure 6 shows the procedure used for the *Grouping step approach* which is also based on the multi-view approach, and table 7 reports the precision, recall, and F1-Score metrics for this approach.

The multi-view learner proposed in these experiments employs the stacked generalization approach, which combines the predictions of each independent view in order to get the final classification result. As shown in figure 4 although the first level learners may have a low accuracy but the combination of their predictions, by generating a new dataset D' and feeding it to train, the meta-learner can improve the accuracy a lot.

We observe an improvement of the accuracy of the overall classification when adding the speed dimension to the ambient air dimensions. We also notice from the confusion matrix in figures

Table 3: Performance of Multi-view Learner (Raw data with/out speed)

class	With Speed			Without Speed		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Street	0.65	0.52	0.57	0.69	0.40	0.50
Bus	0.18	0.12	0.14	0.60	0.33	0.42
Office	0.86	0.93	0.89	0.58	0.43	0.49
Restaurant	1.0	0.22	0.36	1.0	0.10	0.18
Home	0.67	0.65	0.66	0.71	0.87	0.78
Car	0.61	0.81	0.69	0.55	0.75	0.63
Store	0.25	0.05	0.08	0.00	0.00	0.00
Train	0.00	0.00	0.00	0.40	0.08	0.13

Table 4: Performance of Multi-view Learner (Re-sampled data with/out speed)

class	With Speed			Without Speed		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Street	0.72	0.49	0.59	0.59	0.36	0.45
Bus	0.00	0.00	0.00	0.25	0.04	0.06
Office	0.94	0.94	0.94	0.71	0.76	0.73
Restaurant	0.92	0.80	0.86	0.43	0.20	0.27
Home	0.72	0.78	0.75	0.54	0.70	0.61
Car	0.64	0.80	0.71	0.56	0.63	0.59
Store	0.09	0.05	0.06	0.12	0.05	0.07
Train	0.54	0.47	0.50	0.26	0.33	0.29

Table 5: Performance of MLSTM-FCN (Raw data with/out speed)

class	With Speed			Without Speed		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Street	0.55	0.60	0.57	0.53	0.54	0.53
Bus	0.88	0.70	0.78	0.76	0.65	0.70
Office	0.96	0.88	0.92	0.91	0.85	0.88
Restaurant	0.78	0.88	0.82	0.78	0.88	0.82
Home	0.83	0.87	0.85	0.82	0.90	0.86
Car	0.81	0.83	0.82	0.78	0.83	0.80
Store	0.50	0.60	0.55	0.62	0.32	0.42
Train	1.0	0.30	0.46	0.38	0.30	0.33

Table 6: Performance of MLSTM-FCN (Re-sampled data with/out speed)

class	With Speed			Without Speed		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Street	0.43	0.54	0.48	0.50	0.42	0.46
Bus	0.41	0.65	0.50	0.52	0.65	0.58
Office	0.90	0.88	0.89	0.87	0.87	0.87
Restaurant	0.75	0.75	0.75	0.50	0.62	0.56
Home	0.80	0.80	0.80	0.81	0.76	0.78
Car	0.80	0.60	0.69	0.65	0.68	0.67
Store	0.38	0.48	0.42	0.50	0.24	0.32
Train	0.22	0.40	0.29	0.12	0.30	0.17

5a and 5b and the recall and F1 score metrics in table 4 that the model can easily discriminate between the "indoor" and "outdoor" activities, but it cannot perfectly distinguish between the micro-environments inside each category. For example, even though

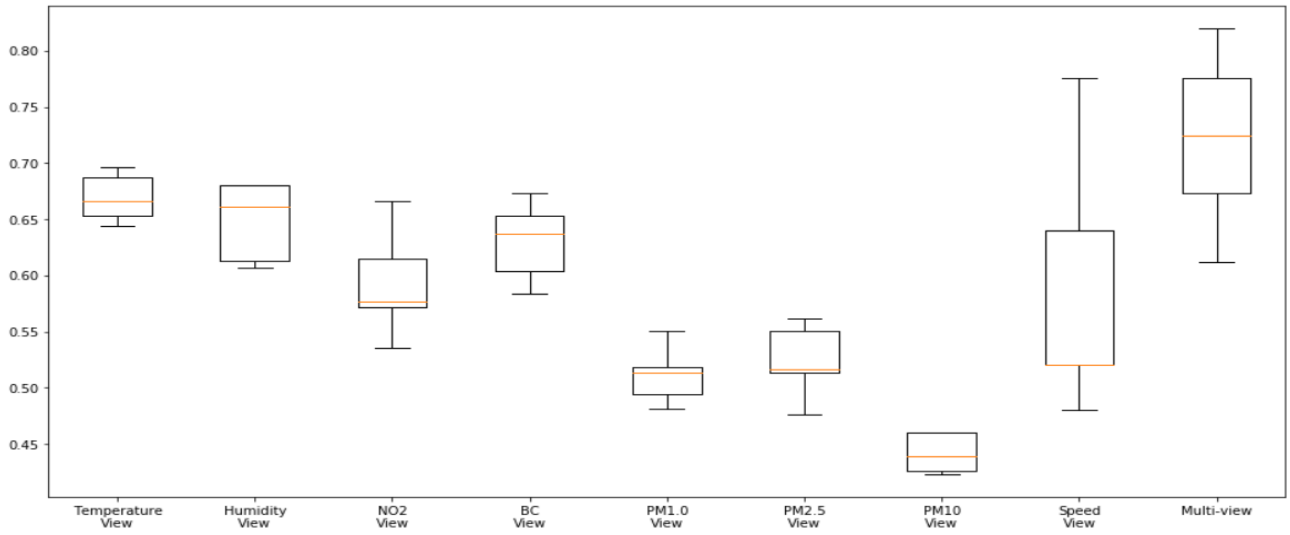


Figure 4: Accuracy among different views (Re-sampled data)

Table 7: Performance of Grouping Step (Re-sampled data with/out speed)

class	With Speed			Without Speed		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Street	0.73	0.59	0.65	0.59	0.35	0.44
Bus	0.60	0.12	0.20	0.00	0.00	0.00
Office	0.92	0.93	0.92	0.80	0.87	0.83
Restaurant	0.91	0.67	0.77	0.20	0.07	0.10
Home	0.86	0.94	0.90	0.75	0.86	0.80
Car	0.71	0.87	0.78	0.66	0.94	0.77
Store	0.46	0.30	0.36	0.40	0.10	0.16
Train	0.29	0.33	0.31	0.33	0.20	0.25

most of the samples in the "Train" micro-environment is falsely predicted as "Car", both "Car" and "Train" micro-environments can be classified as outdoor. Based on this observation, we introduced a grouping step before recognizing the micro-environment. In this step we classify the sample into either an "Indoor" or "Outdoor" environment. Based on the classification result, a model will be specialized for each indoor or outdoor micro-environments. Figure 6 shows the added step and the procedure for the classification.

The accuracy of the classifier in the grouping phase ("Indoor" or "Outdoor") showed a good result when using the resampled data. It reaches 0.82 for data without speed dimension, while for data with the speed dimension it reaches around 0.83. Table7 shows the recall and F1 score for both models trained on resampled data with and without speed. These experiments did not consider the case with original data due to its low performance, in particular, for the minority classes.

5 DISCUSSIONS & PERSPECTIVES

In this section, we discuss the perspectives for improving our multi-view learning model and the possibility for tackling the practical label issue in the context of Polluscope.

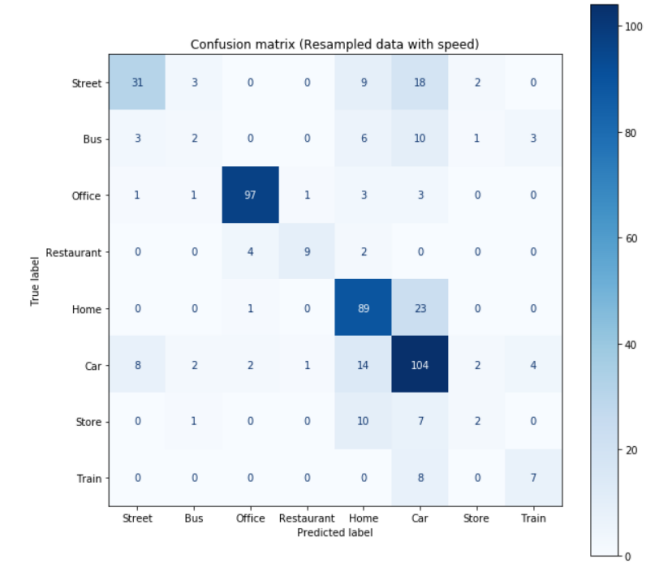
5.1 Multi-view Learner

The multi-view learner adopted in this paper is composed by the base learner (i.e., kNN-DTW) and the meta-learner (i.e., Random Forest), which has greatly improved the performance compared to the single kNN-DTW classifier. The objective of this paper is not to propose the best classifier for MTS classification, but to provide an insight that the multi-view learner is capable of coordinating effectively the information from different variables and achieving more reliable performance than a single base learner. Moreover, the results of the grouping approach which is based on the multi-view approach confirms that there is a clear signature for each micro-environment, thus we can have an effective prediction with this approach.

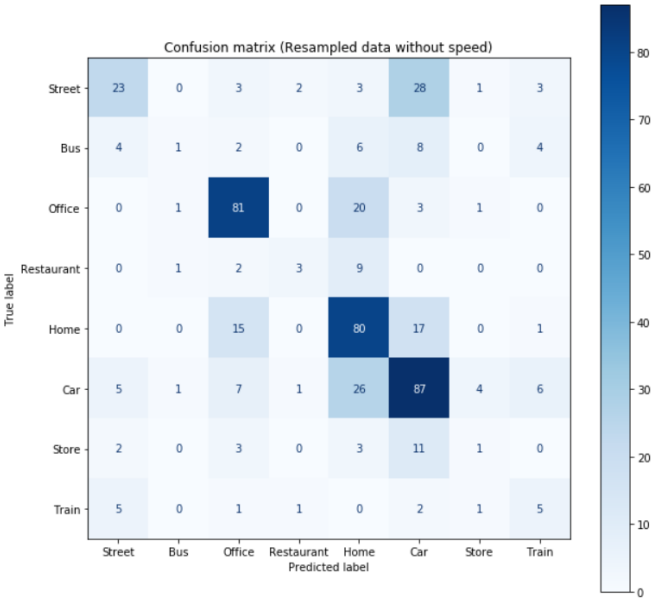
Nevertheless, the kNN-DTW is considered as the baseline for MTS classification and is widely outpaced by the advanced approaches such as Shapelets [27, 32, 33] or the frequent patterns [18]. Essentially, the kNN-DTW captures the global feature based on the distance measure between the entire sequences, while the local features (e.g., the frequent patterns [18], the interval features [5], Shapelets [27], etc.) are more appropriate when a specific pattern characterizes a class. More specifically, a combination of features extracted from different domains may improve dramatically the performance of the base learner [16]. Therefore, one of the perspectives consists in the **optimization** of the base learner and the exploration of the **explainability** of the multi-view learner on both the feature interpretation and the variable importance for building the classifier. The visual representation of Shapelets make them good candidates for such improvement.

5.2 Label Shortage Issue

The label shortage is a practical issue when building the learning model. In particular, in the context of Polluscope, post-labelling for time series sensor data is much more costly than classic data (e.g., image, text, etc.) due to the low interpretability over the real-valued sequence. Therefore, the data need to be annotated during the data collection process. However, certain practical factors limit the availability of labels. For instance, the participants are not always conscious in annotating their micro-environment. Therefore, for certain time periods, no annotations were marked.



(a) With Speed



(b) Without Speed

Figure 5: Confusion Matrix (Re-sampled Data)

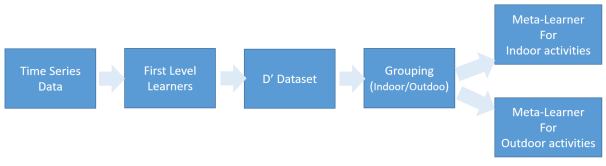


Figure 6: Grouping Process

In order to give an insight for the consistency between the labeled and unlabeled data, and to see if the unlabeled data are valuable for improving the classifier's performance in our context, we conduct a preliminary test on the Polluscope data with the newly proposed semi-supervised MTSC model TapNet [29].

TapNet [29] is a deep learning based approach designed for multivariate time series classification. By adopting the prototypical network [20], TapNet allows learning a low-dimensional embeddings for the input MTS where the unlabelled samples help adjusting the class prototype (i.e., class centroid), which leads to a better classifier than using only the labelled samples. Table 8 shows the semi-supervised learning results on Polluscope data considering or not the *speed* variable. We evaluate the performance of TapNet under different supervision ratios in the training set. The results show that the unlabeled samples and the *speed* variable do improve the performance of the classifier. Besides, the accuracy didn't drop a lot when eliminating the annotations in training set (from ratio=1 for fully labelled to 0.5, and even for 0.2 when only 20% data in labelled), indicating that the collected data within each class is not sparsely distributed, thus learning under weak supervision is reliable with the aid of the unlabeled samples.

Table 8: The accuracy results of TapNet on Polluscope data under different supervision ratios

Condition	Sup_ratio=1	Sup_ratio=0.5	Sup_ratio=0.2
Speed	0.746	0.725	0.717
No speed	0.713	0.703	0.695

Giving the promising results on the data distribution consistency, another avenue worth exploring is to consider and integrate a semi-supervised model into our multi-view learner. Various semi-supervised frameworks are applicable to our model, such as applying self-learning [25] to produce the pseudo labels on the multi-view learner, or adopting the label propagation and manifold regularization techniques [22] on the base learner.

6 CONCLUSION

Activity recognition has gained the interest of many researchers nowadays, due to the widespread use of mobility sensors. Micro-environment recognition is essential in MCS projects such as Polluscope, in order to be able to analyse the individual's exposure to air pollution and to relate it to her context. The major finding of our study is to show to some extent that the ambient air can characterize the micro-environment. Moreover, the accuracy of the model is high enough to consider an automatic detection of the micro-environment without burdening the participants with self-reporting. By using the mobility feature, the accuracy improves slightly though the gain is moderate. Therefore, we can keep characterizing the micro-environment even in the absence of the speed dimension.

We employed different approaches and learners, and conducted a thorough experimental study, which shows the efficiency of MLSTM-FCN and the multi-view approach for time series classification. We have also compared the results with the kNN-DTW classifier which was considered as the baseline.

We have also identified several perspectives of this work, and explored the application of semi-supervised learning to cope with the lack of labels for some classes. In future work, we can use various algorithms for the first level learner and the meta-learner, as multi-view learning is flexible. Finally, we intend to improve the performance of the learned classes by integrating

some a priori rules, like the unlikelihood of being in some micro-environment at some time of day, or of transitions between some micro-environments.

ACKNOWLEDGMENTS

This work has supported by the French National Research Agency (ANR) project Polluscope, funded under the grant agreement ANR-15-CE22-0018, by the H2020 EU GO GREEN ROUTES funded under the research and innovation programme H2020- EU.3.5.2 grant agreement No 869764, and by the DATAIA convergence institute project StreamOps, as part of the Programme d' Investissement d'Avenir, ANR-17-CONV-0003. Part of the equipment was funded by iDEX Paris-Saclay, in the framework of the IRS project ACE-ICSEN, and by the Communauté d'agglomération Versailles Grand Parc – VGP - (www.versaillesgrandparc.fr). We are thankful to VGP (Thomas Bonhoure) for facilitating the campaign. We would like to thank all the members of the Polluscope consortia who contributed in one way or another to this work: Salim Srairi and Jean-Marc Naude (CEREMA) who conducted the campaign; Boris Dessimond and Isabella Annesi-Maesano (Sorbonne University) for their contribution to the campaign; Valerie Gros and Nicolas Bonnaire (LSCE), and Anne Kauffman and Christophe Debert (Airparif) for their contribution in the periodic qualification of the sensors and their active involvement in the project. Finally, we would like to thank the participants for their great effort in carrying the sensors, without whom this work would not be possible.

REFERENCES

[1] Samaneh Aminikhanghahi and Diane J. Cook. 2019. Enhancing activity recognition using CPD-based activity segmentation. *Pervasive and Mobile Computing* 53 (2019), 75 – 89.

[2] Donald J Berndt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *KDD workshop*, Vol. 10. Seattle, WA, USA:, 359–370.

[3] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. 2020. Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges and Opportunities. *arXiv:2001.07416 [cs]* (Jan. 2020). http://arxiv.org/abs/2001.07416 arXiv: 2001.07416.

[4] Heeryon Cho and Sang Min Yoon. 2018. Divide and conquer-based 1D CNN human activity recognition using test data sharpening. *Sensors* 18, 4 (2018), 1055.

[5] Houtao Deng, George Runger, Eugene Tuv, and Martyanov Vladimir. 2013. A time series forest for classification and feature extraction. *Information Sciences* 239 (2013), 142–153.

[6] T. M. T. Do and D. Gatica-Perez. 2014. The Places of Our Lives: Visiting Patterns and Automatic Labeling from Longitudinal Smartphone Data. *IEEE Transactions on Mobile Computing* 13, 3 (March 2014), 638–648. https://doi.org/10.1109/TMC.2013.19

[7] Hafsa El Hafyani. 2020. In *2020 21st IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 246–247.

[8] Hafsa El Hafyani, Karine Zeitouni, Yehia Taher, and Mohammad Abboud. 2020. Leveraging Change Point Detection for Activity Transition Mining in the Context of Environmental Crowdsensing. *The 9th SIGKDD International Workshop on Urban Computing* (2020).

[9] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33, 4 (2019), 917–963.

[10] Hassan Ismail Fawaz, B. Lucas, G. Forestier, Charlotte Pelletier, D. Schmidt, Jonathan Weber, Geoffrey I. Webb, L. Idoumghar, Pierre-Alain Muller, and Franccois Petitjean. 2020. InceptionTime: Finding AlexNet for Time Series Classification. *ArXiv* abs/1909.04939 (2020).

[11] Enrique Garcia-Ceja, Carlos E. Galván-Tejada, and Ramon Brena. 2018. Multi-view stacking for activity recognition with sound and accelerometer data. *Information Fusion* 40 (March 2018), 45–56. https://doi.org/10.1016/j.inffus.2017.06.004

[12] Bin Guo, Zhu Wang, Zhiwen Yu, Yu Wang, Neil Y Yen, Runhe Huang, and Xingshe Zhou. 2015. Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm. *ACM computing surveys (CSUR)* 48, 1 (2015), 1–31.

[13] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Samuel Harford. 2019. Multivariate LSTM-FCNs for time series classification. *Neural Networks* 116 (2019), 237–245.

[14] Baptiste Languille, Valérie Gros, Nicolas Bonnaire, Clément Pommier, Cécile Honoré, Christophe Debert, Laurent Gauvin, Salim Srairi, Isabella Annesi-Maesano, Basile Chaix, et al. 2020. A methodology for the characterization of portable sensors for air quality measure with the goal of deployment in citizen science. *Science of the Total Environment* 708 (2020), 134698.

[15] Sheng Li, Y. Li, and Yun Fu. 2016. Multi-View Time Series Classification: A Discriminative Bilinear Projection Approach. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (2016), 989–998.

[16] Jason Lines, Sarah Taylor, and Anthony Bagnall. 2016. HIVE-COTE: The Hierarchical Vote Collective of Transformation-based Ensembles for Time Series Classification. In *2016 IEEE 16th international conference on data mining (ICDM)*. 1041–1046.

[17] Li Liu, Yuxin Peng, Shu Wang, Ming Liu, and Zigang Huang. 2016. Complex activity recognition using time series pattern dictionary learned from ubiquitous sensors. *Information Sciences* 340-341 (May 2016), 41–57. https://doi.org/10.1016/j.ins.2016.01.020

[18] Guruprasad Nayak, Varun Mithal, Xiaowei Jia, and Vipin Kumar. 2018. Classifying multivariate time series by learning sequence-level discriminative patterns. In *Proceedings of the 2018 SIAM International Conference on Data Mining*. SIAM, 252–260.

[19] Juha Pärkkä, Miikka Ermes, Panu Korpipää, Jani Mäntyjärvi, Johannes Peltola, and Ilkka Korhonen. 2006. Activity classification using realistic data from wearable sensors. *IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society* 10, 1 (Jan. 2006), 119–128. https://doi.org/10.1109/titb.2005.856863

[20] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical Networks for Few-shot Learning. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc., 4077–4087.

[21] Romain Tavenard, Johann Faouzi, Gilles Vandewiele, Felix Divo, Guillaume Androz, Chester Holtz, Marie Payne, Roman Yurchak, Marc Rußwurm, Kushal Kolar, and Eli Woods. 2020. Tslearn, A Machine Learning Toolkit for Time Series Data. *Journal of Machine Learning Research* 21, 118 (2020), 1–6. http://jmlr.org/papers/v21/20-091.html

[22] Jesper E. van Engelen and H. Hoos. 2019. A survey on semi-supervised learning. *Machine Learning* 109 (2019), 373–440.

[23] Baoquan Wang, Tonghai Jiang, Xi Zhou, Bo Ma, Fan Zhao, and Yi Wang. 2020. Time-Series Classification Based on Fusion Features of Sequence and Visualization. *Applied Sciences* 10, 12 (Jan. 2020), 4124. https://doi.org/10.3390/app10124124

[24] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. 2019. Deep Learning for Sensor-based Activity Recognition: A Survey. *Pattern Recognition Letters* 119 (March 2019), 3–11. https://doi.org/10.1016/j.patrec.2018.02.010 arXiv: 1707.03502.

[25] Li Wei and Eamonn Keogh. 2006. Semi-supervised time series classification. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '06)*. Association for Computing Machinery, New York, NY, USA, 748–753. https://doi.org/10.1145/1150402.1150498

[26] David H. Wolpert. 1992. Stacked generalization. *Neural Networks* 5, 2 (1992), 241 – 259. https://doi.org/10.1016/S0893-6080(05)80023-1

[27] Lexiang Ye and Eamonn Keogh. 2009. Time series shapelets: A New Primitive for Data Mining. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09* (2009), 947–956.

[28] Mi Zhang and Alexander A. Sawchuk. 2012. Motion primitive-based human activity recognition using a bag-of-features approach. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI '12)*. Association for Computing Machinery, New York, NY, USA, 631–640. https://doi.org/10.1145/2110363.2110433

[29] Xuchao Zhang, Yifeng Gao, Jessica Lin, and Chang-Tien Lu. 2020. TapNet: Multivariate Time Series Classification with Attentional Prototypical Network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 6845–6852.

[30] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. 2008. Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 312–321. https://doi.org/10.1145/1409635.1409677

[31] Zhi-Hua Zhou. 2012. *Ensemble Methods: Foundations and Algorithms*. CRC press.

[32] Jingwei Zuo, Karine Zeitouni, and Yehia Taher. 2019. Exploring Interpretable Features for Large Time Series with SE4TeC. In *Proc. EDBT*. 606–609.

[33] Jingwei Zuo, Karine Zeitouni, and Yehia Taher. 2019. Incremental and Adaptive Feature Exploration over Time Series Stream. In *2019 IEEE International Conference on Big Data (Big Data)*. 593–602.