# Machine Learning with partially labeled Data for Indoor Outdoor Detection

Illyyne Saffar, Marie-Line Alberi-Morel, Kamal Deep Singh, César Viho

## ▶ To cite this version:

# Machine Learning with partially labeled Data for Indoor Outdoor Detection

Illyyne Saffar
Service Automation,
*Nokia Bell Labs*
Nozay, France
illyyne.saffar@nokia.com

Marie Line Alberi Morel
Service Automation,
*Nokia Bell Labs*
Nozay, France
marie_line.alberi-morel@nokia.com

Kamal Deep Singh
*Laboratoire Hubert Curien,*
University of Saint-Etienne,
Saint-Etienne, France
kamal.singh@univ-st-etienne.fr

Cesar Viho
*IRISA - INRIA,*
University of Rennes 1,
Rennes, France
Cesar.Viho@irisa.fr

*Abstract*—This paper demonstrates the feasibility of an hybrid/semi-supervised classification method for detecting the environment of an active mobile phone, based on both labeled and unlabeled cellular radio data. Precisely, we provide answers to the following question: what is the environment of the mobile user when it is/was experiencing a mobile service/application: indoor or outdoor? Implementing this method within the mobile network is interesting for mobile operators since it has low complexity, is less human intrusive (minimal intervention of mobile users) and more accurate. The semi-supervised classification algorithm learns to identify the environment using large and real collected 3GPP signals measurements. As compared to existing work, in addition to existing parameters used for classification, we propose to also use a radio metric called Timing Advance. It is computed within the mobile network. We empirically validate the innovative semi-supervised algorithm using new real-time radio measurements, with partial ground truth information, gathered daily, weekly, monthly, from indoor and outdoor locations and from multiple typical and diversified environments crossed by mobile users. The study confirms the effectiveness of the proposed scheme compared to the existing supervised classification methods including SVM and Deep Learning.

*Index Terms*—Environment classification, Machine Learning, Indoor Outdoor Detection, 3GPP radio measurement, crowd-sourcing, real user activity.

## I. INTRODUCTION

Recent technological breakthroughs have extended the mobile phones' features, functions and capabilities, which are now used for more than just communicating or affording applications. Recently, mobile devices are being utilized to know the consuming habits of individuals and communities [1], [2], [3]. Thus, our purpose is to inject this learned cognition into mobile 5G networks to help them grow smarter and be more efficient when faced to the increasing complexity of network management combined with numerous new applications and their heterogeneous needs.

As a first step to bring such additional knowledge to the network, we target Indoor/Outdoor Detection (IOD) in this paper. IOD refers to the estimation of the mobile users' environments, that is to infer whether the user is Indoor or Outdoor. IOD is a cornerstone of the user behavior contextualization, which in turn can be used for learning the user behavior, adapting mobile network resources, etc [4], [5]. The idea is to have more information on the user like knowing his environment type or his location.

IOD can be performed automatically and in real-time using machine learning techniques, which in turn need data for learning. Thus, data collection is the first phase of designing IOD solution based on machine learning. Recently, a new crowd-sourcing approach [6], [7] is becoming popular for collecting and analyzing real and large network measurement datasets coming from mobile phones or any other connected devices. This method exploits smartphones (with built-in cellular network interface) with their various measurement sensors. Additionally, data obtained from smartphones has the natural mobility vector of people carrying them. This ensures cost-effective, continual and fine-grained spatio-temporal monitoring and analyses of mobile networks. For our work, we propose to investigate this concept of large and real crowd-sourced measurements for IOD. We also propose to extend it to mobile networks to deal with the challenge of detecting the environmental context of mobile users from network side. The idea is to collect data, which is measured or derived within network, and then consider it as an input for the machine learning based classifier used for training, learning and then detection. The data measured by multiple UEs during their connection is sent to eNB, using standardized procedures.

Such solutions are interesting for mobile network operators that wish to exploit cognition of user behavior to optimize/customize their service delivery with minimal intervention of the users. Furthermore, such measurements, as an alternative to coverage modeling or drive tests [6], capture reality well, reveal real life of a mobile user while at the same time being less expensive. This method can then be implemented by the operators in their networks, as a generic solution, independent of the implementations of particular manufacturers. Consequently, it allows the mobile network to exploit direct measurements at user side to deduce contextual factors such as the user environment.

In 4G/5G cellular networks, such solutions are technically feasible since enormous amount of mobile measurement data is collected by the mobile terminal. This data is regularly sent to the network using standardized protocols and interfaces during each UE's connection to the cell (on a per-procedure basis and on a network defined event basis). This measurement data is referred to as LTE UE Measurement Data (LUMD) [8]. LUMD contains rich information on mobile performance and RF metrics such as signal strength (Reference Signal Receive

Power or RSRP), signal plus interference and noise strength (Reference Signal Receive quality or RSRQ). It also includes the Channel Quality Indicator (CQI) that is a function of SINR.

In this work, we aim to achieve the following objectives:

- (1) infer the user environmental context, from certain LUMD metrics collected in crowdsourcing mode and the radio metric, Timing Advance, assessed by the network when the user is connected to a session. In fact, the environment considered is divided into two main types:
  - Indoor: at home, in restaurant, in cafe/ at work or in other building types, etc.
  - Outdoor: pedestrian, running or in car moving with high speed.
- (2) consider the constraint that the inference shall be done at network side with minimal human interaction or intervention.

To achieve (1) and (2), we design a method for training IOD automatic classifier based on a weakly or partially labeled crowdsourced dataset. Such dataset reduces human intervention to the lowest possible. Indeed, the labeled data, used for machine learning training, is either tagged manually or automatically. Manual data tagging can be expensive, complex and even unfeasible for mobile operators if they have to tag all the collected crowdsourced data.

In this paper, we are interested in Machine Learning (ML), one of the popular techniques, for automatic IOD. Among ML families, we consider supervised learning and more particularly semi-supervised learning which can be seen as a mix of supervised and unsupervised approaches. Supervised learning is more adapted for classification tasks. It uses labeled data to learn the mapping between data and the labels. Unsupervised learning looks for patterns and structures within the data for tasks such as clustering. The semi-supervised learning, which is an hybrid approach, is becoming popular with growing abundance of data in this era. It proposes a learning scheme based on partially or weakly labeled dataset in order to achieve a classification task or a function approximation task. In our case, semi-supervised learning allows the mobile operator to use labeled data from a few users and combine it with lot of unlabeled and easily available data collected from several users. This combination allows to learn all possible environment types related to the user behavior.

The rest of this paper is organized as follows. Section II describes the main IOD works in literature. In Section III, a comparative analysis of crowdsourcing and drive-test data collection modes is provided. In section IV, results with supervised classification and clustering algorithms are given. Section V and VI present a new Deep Learning-based semi-supervised learning approach proposed for IOD from the network side. Section VI discuss the results.

## II. RELATED WORK

In the literature, the IOD issue has not been largely studied: only few works address it. Proposed solutions are usually divided in to two categories [9]. IOD is either considered as a statistical issue where a weighted score or a threshold is defined to determine the mobile environment, or as a classification problem sorting mobile users between multiple classes. In most of these works, only two classes are considered (Indoor/Outdoor) but, in some works, three classes are selected (e.g. Indoor/Semi-Outdoor/Outdoor). The Figure 1 shows an illustration of the whole dependency of existing classes.
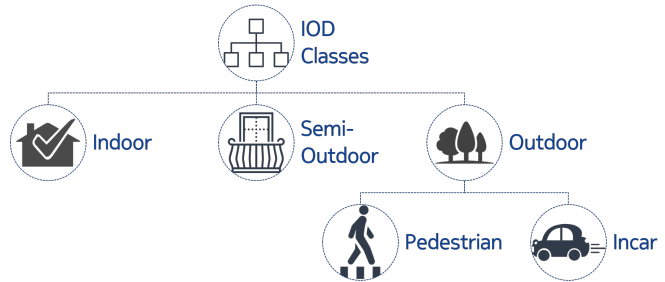


Fig. 1. Example of IOD classification scheme: in 3 main classes

In addition to such categorization, IOD problem can also be distinguished based on the location where IOD is performed, either at the mobile terminal side or at the mobile network side. In the following, we highlight some of the works dealing with the IOD issue, presenting them according to this classification.

In first category, [10] looks at a threshold of signals collected from some phone sensors related to: radio signals, cell signal strength, light intensity as well as the magnetic sensor to infer whether the mobile user is indoor or outdoor. However, this threshold is specific to the experimental settings where it is calculated. It is not generalizable to new environments. Thus, using just a threshold decreases the IOD accuracy. Similar to [10], the work in [5] also uses the same signals, but also considered sound intensity, battery temperature and the proximity sensor. For IOD, they propose a semi-supervised approach: a co-training solution. They use 2 classifiers in parallel with a weighted score of classification probability to improve the final performance of IOD. For every classifier, they select a different set of sensors to learn different perspectives and patterns. This work shows high performance (more than $90\%$ of accuracy) in the detection of new instances in unknown environments. However, the impact of this work is limited since their database is not highly representative. Indeed, the used data set was only collected in three places (the campus area, city center, residential area) which are not enough to train a general IOD system.

The work in [4], proposes a video streaming optimization based on adaptation as a function of the user location in time. For that, IOD is computed via a Bayesian detector that combines measurements from two smartphone sensors to decide the user environment type.

In second category, in [11] authors optimize the use of radio measurements in wireless networks. Literally, they use radio signal measurements collected in different situations

of mobility with varying speed (low, medium, high) namely (pedestrian, incar and unmoving). They dynamically estimate the signal attenuation. This in turn helps them to efficiently classify mobile user environment (pedestrian, incar, unmoving) and finally improves the handover process. Authors assume that once the signal power attenuation is estimated correctly, we can easily come to classify whether the mobile user is pedestrian, in car or unmoving. This is because the measured power signal for an unmoving user does not show too much variations unlike the incar or pedestrian cases. Nevertheless, this proposition is still at an early stage and it has not been thoroughly developed yet. In [8], the main issue is to localize the mobile user by estimating its longitude and latitude in a most possible accurate way. For this, they made the assumption that mobile users are outdoor, thus giving rise to the importance of IOD and the necessity to classify the user environment. For the classification task, they used RSRP and RSRQ signals and tested many algorithms: SVM, logistic regression and random forest. SVM was the retained solution since it performed best.

In this paper, we focus on the IOD automation within the network side using machine learning algorithms. They are trained using large real dataset while minimizing the mobile user interaction (minimal labels). We look at the performance in terms of $F1 - scores$ of supervised and semi-supervised IOD methods. Goal is to evaluate the minimal amount of labeled data required for obtaining good IOD performance.

## III. Collected Data for IOD

In this section, we analyze the statistical differences by focusing on the empirical cumulative distribution function (CDF) between indoor and outdoor environments, using a large and real data-set collected at multiple places, many environments. We illustrate the impact of the two environments on the empirical CDFs, according to where the data is collected.

### A. Data Description

Our large data set consists in Time, 3 LUMD radio signals, the metric Timing Advance (TA) and the label when it is known. Thus, it has a vector of 6 features with the label:

- Time: time of signal record
- RSRP: the average received power of the Reference Signal (RS) between -140 dBm to -44 dBm [12], sent by eNB.
- RSRQ: the ratio between RSRP and RSSI (Received Signal Strength Indicator) between -19.5dB and -3dB [12], that represents the total power of the received signal (including the transmitted signal, the noise and the interference).
- CQI: indicator reported by UE to eNB that gives the most appropriate modulation scheme and coding scheme to be used for transmission [13].
- TA: used to control Uplink signal transmission timing. It is indicated by eNB to UE via a Timing Advance command [14].

The set of these signals has been collected during 9 months, $24h/7$ (From October 2017 until June 2018), with an average of 1 measurement per 15 seconds while the mobile phone session is active and 1 measurement per 2 minutes otherwise.

The dataset is made of $40\%$ of labeled data and $60\%$ of unlabelled data. The 9 months collection has been performed in many different environments like mountain, beach, forest, companies, cafes, streets, bars, parks, restaurants, lakes, etc... It was also performed in many cities and places like country-side, villages, small cities, metropolis, and different countries, but for this paper we are only studying the data collected in France (Figure 2). This long collection period allows us to have data reflecting all weather types: Heavy Rain, Foggy, Sunny, Snowy, Windy, Rainy,... i.e. almost the 4 seasons. Therefore with this campaign of data measurement we try to be as close as possible to the complexity and the variety of a mobile user moving in real world.
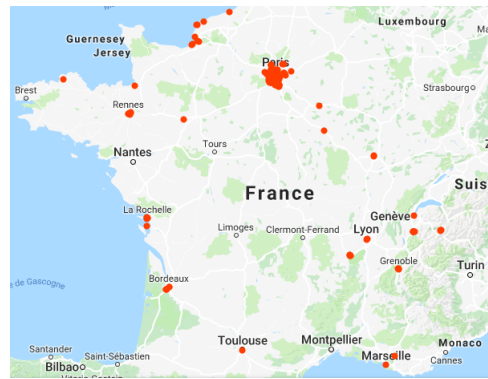


Fig. 2. Data collection Points in France: multiple environments and places

### B. Data collection: crowdsourcing vs. drive-test mode

In crowdsourcing mode, the collected data consists of signals measured by the mobile phone and sent to the eNB. Our dataset described in the previous subsection has been collected using this mode. Figures 3 shows the empirical cumulative distribution functions (CDFs) of RSRP and CQI obtained with the dataset. The significant offset between the indoor and the outdoor curves, results from substantial difference and attenuation variation in radio signal propagation. It is mainly due to reflection, diffraction, dispersion and attenuation experienced in indoor environment. However, we note that there is some overlap between the ranges of RSRP and CQI values. Also the extreme values seen in the two indoor and outdoor CDFs (located in tails) get similar and the division between the two gets blurred. The behaviour at the juncture of extreme values can be explained by the ambiguous characteristics of the environment when a user is at high speed (Train, car...) or when he is in a semi indoor environments (like balconies, semi-open building, near a window.., etc. We argue that these points are ambiguous and will pose a good challenge for supervised classification, since they can be indifferently classed indoor or outdoor at the same time.
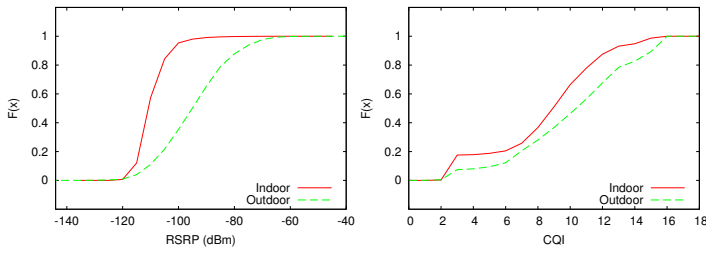
Fig. 3. Empirical CDF for measured RSRP (left) and CQI (right) in crowdsourcing mode: multiple environments and places - Indoor (red) and Outddor (green).
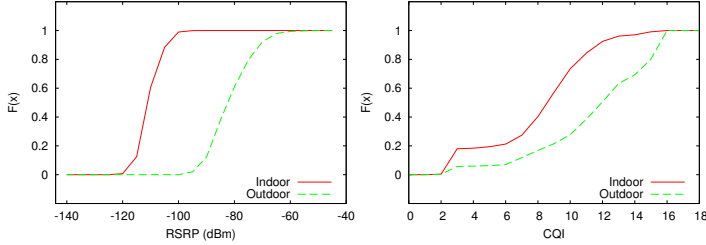


Fig. 4. Empirical CDF for measured RSRP (left) and CQI (right) in drive-test type mode: specific environments and places - Indoor (red) and Outddor (green).

An alternate data collection mode, widely used to collect data, is the drive-test mode. However, this mode imposes limits on capturing the reality through the data collected in this mode. Such data collection campaigns are run for limited hours per day during short period (couple of weeks) and at some specific places. To model this way of collecting data, referred as drive-test mode, we extract a portion data (EPD) from the whole dataset. We aimed by this selected EPD data to be as close as possible to the type of places where the drive-test was performed by one of the top 3 American operators in New York City in [8]. Therefore to build EPD we consider data only in metropolis (Paris and southern suburbs see figure 5). Indeed, Paris as metropolis, has a dense and specific architecture which allows better comparison with NYC. Concerning indoor data, we selected instances where the user was strictly indoor and, thus, not in "semi-indoor" positions like semi-open building or balconies,...etc. For outdoor data, we chose the instances where the user was either pedestrian or in vehicle in different city streets (limited speed). Thus, to mimic drive-test we consequently ignored data coming from environments like subway/ countryside/ forest/ beaches/ Mountains/ .../ etc. We did this to enable a fair comparison between the two modes. Figure 4 shows well separated RSRP empirical cdfs between the classes indoor and outdoor. The superimposed points of both the cdfs we judge conflicting have disappeared. The overlap between both the cdfs, which previously led to ambiguity, has disappeared. This is due to the significant distance between the indoor and the outdoor curves. In the case of CQI cdfs we notice a similar phenomenon. This analysis allows us to argue that supervised classification will run better

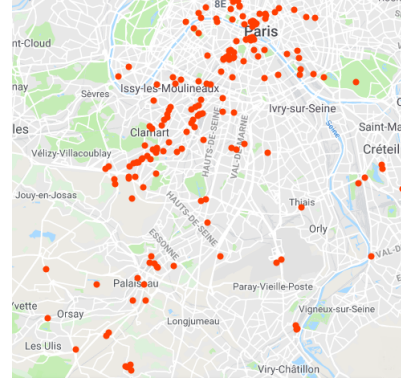on labeled dataset collected in drive-test mode as compared to obtained through crowdsourcing mode.



Fig. 5. The Data collection Points of EPD in drive-test like mode: Paris and southern suburbs

## IV. CLASSIFICATION USING SUPERVISED LEARNING OR CLUSTERING

After analyzing the statistical properties of I/O environments, we first evaluate the accuracy and the performance of supervised classifiers for IOD. For this, we use the accuracy metric which is the ratio of correctly classified instances divided by the total instances and the metric $F1 - score$ that is by definition the weighted average of Precision and Recall according to the following relation:

$$F1 - score = 2.\frac{Precision.Recall}{Precision + Recall}$$

where precision is the number of correct positive results divided by the number of all positive results returned by the classifier, and recall is the number of correct positive results divided by the number of all relevant samples. $F1 - score$ is one of the most used metrics in case of unbalanced data classes. Indeed, the statistics of our data show that the data proportion between indoor and outdoor classes is unbalanced $65\%$ Indoor vs. $35\%$ Outdoor. This reflects the reality since people, in general, spend more time at home or in indoor environments than in outdoor environments. For the experiments, we divided the dataset as follows: $70\%$ for training, $30\%$ for validation and test. We evaluate the impact of both input pairs (RSRP, RSRQ), which is the reference input for IOD in the literature, vs. (RSRP, CQI), in three cases:

- Training and evaluation on labeled EPD collected in drive-test like mode (see Table I),
- Training on labeled EPD and evaluation on the rest of the labeled data of crowdsourcing mode, thus operating with unknown environments (see Table II) and,
- Training and evaluation on labeled data collected in crowdsourcing mode (see Table III).

As shown in the table I, running either classification (SVM, Random Forest, Neural Network) or clustering ($k$-means) algorithms on EPD, obtained from drive-test like mode, shows an excellent performance with an $F1 - score$ of 99%, which is

close to the reference result found in literature [8]. However, when the algorithm trained on EPD is used to perform IOD directly on crowdsourced data, a dramatic performance deterioration is observed as seen in table II. The best algorithm is SVM, which gives an $F1 - score$ of 61.7%. But, this is not an acceptable performance for IOD. In third case, the performance of supervised classifier where training as well as evaluation is performed on the crowdsourced labeled data, is shown only for the case of SVM, which performed best. Table III shows a noticeable enhancement of $F1 - score$ to 83.71%. This is a moderately acceptable performance. We are still far from the reference in the literature. For the target, we can assume that an error of $5 - 8\%$ is tolerable for the IOD system. Indeed, while dimensioning of mobile networks, an error up to $10\%$ is qualified as an admissible error rate. Additionally, Tables I, and III show that (RSRP, CQI) as input provides similar results as (RSRP, RSRQ) when used for classifying EPD or crowdsourcing data. The results are even slightly better in case of table II with (RSRP, CQI). RSRQ and SINR (note that CQI is based on SINR) are both radio measurements that depend on signal and interference strength. The results shows that the information contained in CQI is also useful for IOD and thus, (RSRP, CQI) is also a good candidate for IOD.

| Algorithm | RSRP-RSRQ | | RSRP-CQI | |
|---|---|---|---|---|
| | Accuracy | F1-Score | Accuracy | F1-score |
| $k-$means | 99,68% | 99.48% | 99.67% | 99.47% |
| SVM | 99.75% | 99.59% | 99.76% | 99.60% |
| NeuronalNetwork | 99.50% | 99.18% | 99.57% | 99.28% |
| RandomForest | 99.83% | 99.72% | 99.77% | 99.62% |

TABLE I
CLUSTERING AND CLASSIFICATION PERFORMANCE: TRAINING AND
EVALUATION ON LABELED DATA (EPD) OF DRIVE-TEST LIKE MODE

| Algorithm | RSRP-RSRQ | | RSRP-CQI | |
|---|---|---|---|---|
| | Accuracy | F1-Score | Accuracy | F1-score |
| $k-$means | 61.41% | 60.07% | 59.64% | 57.77% |
| SVM | 56.56% | 36.13% | 62.69% | 61.71% |
| NeuronalNetwork | 50.90% | 44.58% | 62.55% | 61.54% |
| RandomForest | 62.93% | 61.99% | 62.63% | 61.59% |

TABLE II
CLUSTERING AND CLASSIFICATION PERFORMANCE: TRAINING ON EPD
AND EVALUATION ON LABELED DATA OF CROWDSOURCING MODE

As we guessed, the performance of IOD classification when

| Algo. | RSRP-RSRQ | | RSRP-CQI | | RSRP-CQI-TA | |
|---|---|---|---|---|---|---|
| | Accur. | F1-S. | Accur. | F1-S. | Accur. | F1-S. |
| SVM | 85.48% | 83.66% | 85.54% | 83.71% | 90.17% | 89.11% |

TABLE III
SVM PERFORMANCE: TRAINING AND EVALUATION ON LABELED DATA OF
CROWDSOURCING MODE

trained only on EPD and then tested on the crowdsourced data drops in terms of $F1 - score$ and accuracy. This is due to the presence of ambiguous points combined with unknown environments not included in the drive-test data. Consequently,

learning the user environment, only based on drive-test data, is thus not enough to learn the complexities of users' real life.

We continue the study with SVM as the reference supervised classifier. The inputs (RSRP, CQI) are more appropriate for doing IOD from network infrastructure point of view since these signals are sent more regularly to eNB than RSRQ. Lastly, we propose to add a new signal, called Timing Advance (TA), to enhance the IOD performance. The idea is to exploit the information of distance between eNB and the mobile users embedded in TA parameter. This would help the supervised classifier to classify the ambiguous points (e.g. measurement points with low RSRP, but close to eNB, etc.). So, in case of (RSRP, CQI, TA), the SVM performance reaches an $F1 - score$ of 89.11% and an $Accuracy$ of 90.17% (Table III). As a result, with the addition of TA, IOD using SVM performs better leading to a gain of 6%. TA notably contributes to solve the classification issue of some ambiguous points.

## V. HYBRID/SEMI-SUPERVISED APPROACH

To avoid performance degradation when facing new unknown environments, it is preferable to train the IOD classifier using a more diversified dataset. From a data collection point of view, it is more of an interest for the operator to collect massive partially tagged data. Indeed, first, during online labelling it alleviates the network charge by limiting the amount of UL signalling (all labels) sent to eNB and, second, reduces the complexity and the time for tagging data. Thus, the idea is to use the available tagged data, which is costly to obtain, and combine it with untagged data, which is easy to obtain, for classifier training. However, one of the main questions is: how much percentage of tagged data is needed for satisfactory performance of the intelligent IOD system?

We suggest a semi-supervised learning system (HSSL) that can learn additional new environments, without the need to involve more users to gather the ground truth (the indoor/outdoor tag). As in [5], [15], [16], our approach uses both tagged and untagged data in order to improve the IOD classifier training, while maintaining the same good performances for a given ratio of tagged and untagged data. The proposed system is composed of 2 main modules as shown in Figure 6. The role of first module is to label the untagged data. It uses a unsupervised clustering algorithm, called "Bayesian Gaussian Mixture" (BGM) which is fast and efficient. The second module uses this tagged data output to learn the user environment via a supervised learning classifier that can be SVM or also Deep Learning algorithm.

Recently Deep Learning (DL) approaches have emerged which show improvements as compared to classical approaches such as SVM [17], [18]. Moreover, from an operator point of view, IOD is a complicated task since millions of users are considered during a longer period which heavily increases the dataset size. Therefore we propose to also investigate Deep Learning (DL) over the huge crowdsourced dataset. After configuring DL using a Grid Search to find the best parameters that best optimize the IOD classifier, we conduct

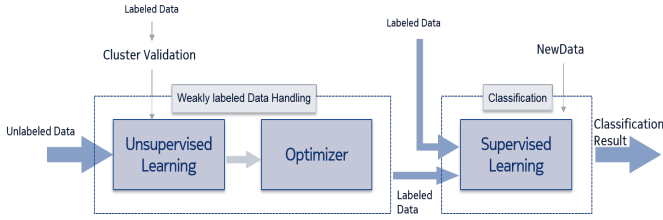a comparative study between SVM, HSSL (using SVM), DL and HSSL (using DL).



Fig. 6. IOD Learning Scheme for Weakly Labelled data: an hybrid/Semi-supervised Machine Learning approach

In this approach, the first module detects 2 clusters. Once detected, they are employed to label the untagged data. An optimizer module then processes the data before sending it to the second module. It corrects and minimizes the labeling errors resulting from clustering. For this, we assume that a user can not change his environment twice in 30 seconds. The idea can be explained from the following example. Imagine that we have three consecutive points, very near in time, in the dataset. If, for example, the first point is mapped as indoor, the second is mapped as outdoor and the next point, very near in time, is again mapped as indoor, then we assume that there is an error in mapping. This is because a user cannot change its environment two times so quickly. Thus, the optimizer module detects and corrects such errors.

Let $E_t$ the environment type of the user at the moment $t$ and the different measurement times $t-1$,$t$ and $t+1$. If the difference between $(t-1, t+1)$ is equal or less than $60s$ then $E_{t-1} = E_t = E_{t+1}$. The optimizer parses the data tagged with the cluster verifying this assumption and then correcting the BGM prediction if necessary, see Algorithm 1. The clustering

---

**Data:** output of the cluster : tagged data
**Result:** Optimization and tags correction
**for** $E_t$ *in clustering_Tagged_DataSet* **do**
    **if** *Diff(t − 1,t + 1)$\leq$ 60s* **then**
        **if** $E_{t-1} = E_{t+1}$ *And* $E_{t-1} \neq E_t$ **then**
            $E_t \leftarrow E_{t+1}$
        **end**
    **end**
**end**

**Algorithm 1:** Time Optimizer

---

and the optimizer together form the first module of the HSSL system which deals with labeling of the unknown data tags. The input of first module consists of untagged data. The output is considered as the first input of the second module dealing with the supervised classifier. This output is a vector of 4 measurements [RSRP, CQI, TA, Class*], where Class* is the estimated labels by the first module. The second input of the second module is composed of the labels (the ground truth) forming a measurement vector of size 4 [RSRP, CQI, TA, Class].

## VI. RESULTS AND DISCUSSION

This section evaluates the performance of HSSL on the crowdsourced data. It provides an answer to the question on what is the ideal amount of tagged data required with the untagged data so that the performance of the intelligent IOD system achieves a $F1 - score$ higher than $90\%$.

We have used both scikit-learn [20] and keras [21] in python for the HSSL implementation. The DL module is a feed forward neuronal network (fully connected) with 8 hidden Layers using ReLU as the activation function. Actually, ReLU is the most widely used activation function while designing neural networks today. The main advantage of using ReLU over other activation functions is that it does not activate all the neurons at the same time. It leads to a sparse network that is efficient and easy for computation. As for the last layer (the output layer) we used a sigmoid activation function since we look for a binary classification either 0 or 1 (for indoor/outdoor environments).

The HSSL evaluation is first done in 2 validation steps:

- (i) The performance of first module (BGM + Optimizer) provides an $F1 - score$ of $85.99\%$.
- (ii) The performance of second module (supervised learning using SVM or DL) shows an $F1 - score$ of $89.11\%$ with SVM and of $92.81\%$ with DL.

Once confirmed that both modules have convincing performance, we evaluate the whole HSSL system. The system receives both labeled and unlabeled data as inputs. The evaluation goal is to find out for what percentage of labeled data (and unlabeled data), the performance of HSSL goes above the target $F1 - score$ of $90\%$. For this, we aim to compare HSSL (including SVM or DL) with SVM and DL, alone, when trained over same amount of tagged data (with the only difference that HSSL in addition also uses untagged data). Figure 7 shows performance of HSSL (DL or SVM) and of supervised SVM and DL for different percentages of labeled data. We observe that IOD performs better using DL than using SVM in both cases. We also observe that only DL and HSSL(DL) achieve both the tolerable error of $5-8\%$ for IOD system in mobile network. However, we note that HSSL is slightly better than DL for almost all percentages of labeled data. HSSL(DL) reach the maximal $F1 - score$ of $93\%$ for the distribution of $65\%$ labeled data and $35\%$ unlabeled data.

Also, it can be seen that for an operator a dataset composed only of $10\%$ of tagged data, approx. 1 month of collected data out of total 9 months, is enough to learn the user environment.

To conclude, the proposed HSSL system trained with a partially tagged data set, is able to make a good distinction of the user environment. We also note that supervised DL is better than SVM. This is because DL uses several layers of neurons and is able to capture more mappings. HSSL(DL) showed only slightly better performance as compared to DL. This is because we studied simple IOD with detection of only 2 classes. In future, we plan to study detection of more environments such as in-car, pedestrian, etc., instead of just outdoor or indoor. We will compare HSSL(DL) with DL with
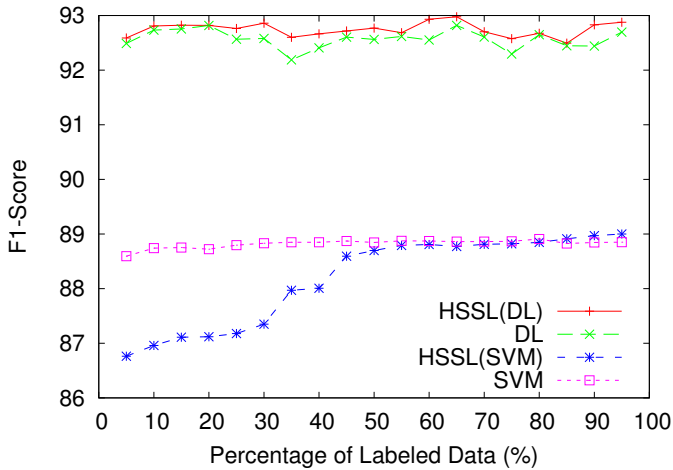
Fig. 7. Evaluation of the HSSL System: $F1 - score$ vs. percent of labeled data.

more classes, environments and yet more users and data. The hypothesis will be that adding unlabeled data might improve the HSSL performance more.

Nevertheless, HSSL with SVM trains faster with a duration of $32.95s$ on a machine having 12 CPUs and 32 Go of RAM. Training with DL is slower: about 23.35 minutes using the same machine. SVM and HSSL with SVM converge quicker than DL. In future, we will try training using a GPU.

## VII. CONCLUSION

In this paper, we investigated the problem of IOD performed at network side using 3GPP signals and Timing Advance data collected inside the infrastructure. We first showed that using a drive test dataset is insufficient to mimic the real world complexity and reveal the real user behavior. By diversifying the environments more (using a highly representative crowd-sourced dataset) during the training phase, we showed that the more environments we have for the training phase, the better the supervised classifier performs. We also showed that adding a new parameter, Timing Advance, can improve IOD performance.

To address the fundamental issue of the model adaptation to new and diversified environments without making it hard and expensive for the operators (specially due to the labelling task) we proposed a new hybrid/semi supervised learning (HSSL) system. The HSSL system presents satisfactory performance even when facing unknown environments.

We plan to extend our work on IOD in future and address the IOD issue by considering systems that take the time variations into account. Thus, probably using other algorithms with time correlation, like LSTM, would boost the HSSL system and would probably decrease the required portion of labeled data to obtain $F1 - score$ of $95\%$.

## ACKNOWLEDGMENT

## REFERENCES

[1] BULUT, E., and SZYMANSKI, B. K., *Understanding User Behavior via Mobile Data Analysis*, Proc. IEEE ICC Workshops, Dynamic Social Networks, DYSON, London, June 8, 2015, pp. 1548-1553

[2] B. O. Holzbauer, B. K. Szymanski, and E. Bulut, *Impact of Socially Based Demand on the Efficiency of Caching Strategy*, in Proceedings of IEEE PerCom 2014, IQ2S workshop, Budapest, Hungary, 2014.

[3] Q. Xu, Z. M. Mao, A. Arbor, J. Erman, F. Park, A. Gerber, J. Pang, S. Venkataraman, *Identifying Diverse Usage Behaviors of Smartphone Apps*, Proceedings of the 2011 ACM SIGCOMM conference, Internet measurement conference, 2011, pp. 329-344.

[4] MEKKI, Sami, KARAGKIOULES, Theodoros, et VALENTIN, Stefan. HTTP adaptive streaming with indoors-outdoors detection in mobile networks. arXiv preprint arXiv:1705.08809, 2017.

[5] RADU, Valentin, KATSIKOULI, Panagiota, SARKAR, Rik, et al. A semi-supervised learning approach for robust indoor-outdoor detection with smartphones. In : Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems. ACM, 2014. p. 280-294.

[6] MARINA, Mahesh K., RADU, Valentin, et BALAMPEKOS, Konstanti-nos. Impact of indoor-outdoor context on crowdsourcing based mobile coverage analysis. In : Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges. ACM, 2015. p. 45-50.

[7] CAINEY, Joe, GILL, Brendan, JOHNSTON, Samuel, et al. Modelling download throughput of LTE networks. In : Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on. IEEE, 2014. p. 623-628.

[8] RAY, Avik, DEB, Supratim, et MONOGIOUDIS, Pantelis. Localization of LTE measurement records with missing information. In : Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE Inter-national Conference on. IEEE, 2016. p. 1-9.

[9] EDELEV, Sviatoslav, PRASAD, Sunaina Nelamane, KARNAL, He-manth, et al. Knowledge-assisted location-adaptive technique for indoor-outdoor detection in e-learning. In : Pervasive Computing and Com-munication Workshops (PerCom Workshops), 2015 IEEE International Conference on. IEEE, 2015. p. 8-13.

[10] ZHOU, Pengfei, ZHENG, Yuanqing, LI, Zhenjiang, et al. Iodetector: A generic service for indoor outdoor detection. In : Proceedings of the 10th acm conference on embedded network sensor systems. ACM, 2012. p. 113-126.

[11] ALAYA-FEKI, Afef Ben Hadj, LE CORNEC, Alain, et MOULINES, Eric. Optimization of Radio Measurements Exploitation in Wireless Mobile Networks. JCM, 2007, vol. 2, no 7, p. 59-67.

[12] 3GPP TS 3GPP TS 36.133: "Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management", Release 8.

[13] 3GPP TS 3GPP TS 36.213: "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures", Release 8.

[14] 3GPP TS 3GPP TS 36.321: "Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification", Release 8.

[15] ZHU, Xiaojin et GOLDBERG, Andrew B. Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning, 2009, vol. 3, no 1, p. 1-130.

[16] ZHANG, Jun, CHEN, Xiao, XIANG, Yang, et al. Robust network traffic classification. IEEE/ACM Transactions on Networking (TON), 2015, vol. 23, no 4, p. 1257-1270.

[17] LECUN, Yann, BENGIO, Yoshua, et HINTON, Geoffrey. Deep learning. nature, 2015, vol. 521, no 7553, p. 436.

[18] GOODFELLOW, Ian, BENGIO, Yoshua, COURVILLE, Aaron, et al. Deep learning. Cambridge : MIT press, 2016.

[19] JIANG, Chunxiao, ZHANG, Haijun, REN, Yong, et al. Machine learning paradigms for next-generation wireless networks. IEEE Wireless Com-munications, 2017, vol. 24, no 2, p. 98-105.

[20] PEDREGOSA, Fabian, VAROQUAUX, Gal, GRAMFORT, Alexandre, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research, 2011, vol. 12, no Oct, p. 2825-2830.

[21] CHOLLET, Franois, et al. Keras. 2015.