Deep Learning based Speed Profiling for Mobile Users in 5G Cellular Networks.

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

Deep Learning based Speed Profiling for Mobile Users in 5G Cellular Networks

Illyyne Saffar*‡, Marie Line Alberi Morel*, Kamal Deep Singh†, Cesar Viho‡

*Nokia Bell Labs, Nozay, France, ‡Univ Rennes, Inria, CNRS, IRISA, France

†Univ Lyon, UJM Saint-Etienne, CNRS, Laboratoire Hubert Curien UMR 5516, Saint-Etienne, France

Email: {illyyne.saffar,marie_line.alberi-morel}@nokia.com, kamal.singh@univ-st-etienne.fr, cesar.viho@irisa.fr

***Abstract*—Future mobile networks promise to be more intelligent and to guarantee a better service for users. This intelligence can be highly accentuated by cognition of mobile users' behavior and conditions. The speed is an important element of user profile. We are interested in real time speed profiling by detecting speed range of an active user. We refer to this as Mobile Speed Profiling (MSP) of users. Indeed, performing MSP can notably improve the self-adaptation and self-optimization capabilities of these networks. It can help mobile network in resource management and handover management. This paper introduces a Deep Learning based solution to automatically construct the MSP of a mobile user. We empirically evaluate the effectiveness of our approach using real-time and highly representative radio data. This data includes ground truth information and the whole dataset has been gathered massively from many diversified mobility situations. Results show that the profiling of UE's speed with fine granularity or on multiple ranges can be achieved with high accuracy on real data measured in heterogeneous deployments for 5G networks.**

***Index Terms*—LTE/5G, Mobility Speed Profiling, Deep Learning, Multi-output Classification, 3GPP radio measurement, crowdsourcing data, real user activity.**

I. Introduction

New 5G networks are promising many new technologies and innovations. They are designed to ensure that "Every thing is connected to every thing at any time". Actually, we are interested in one of the fundamental innovations of the 5G wireless networks, which is the context-aware service delivery. Such innovation will allow operators to optimize every connection for every user, device and application, by shifting the network planning, and monitoring tools from a network-centric to a user-centric view.

The use context awareness is a prerequisite for inferring the mobile user behavior and thus, precisely profiling the mobile users. The context of use, namely the context of mobile user, is defined in [ISO 13407:1999] standard as the characteristics of the users, tasks and the environment in which the system is used. Consequently, defining all the usage situations of mobile services leads to addressing of 6 main questions according to Kipling's method (who/ where/ what/ when/ why/ how) [1]. This paper rather focuses on the 'How' question related to the user mobility that's to say: "How is he consuming his application (moving or static)?" or "A user prefers to use his phone while being static or in mobility? What is his speed range while being connected and using phone?"

Users move, changing locations, around 3, 4 or more times per day. This daily routine results in different speed profiles of mobile users. Analyzing and estimating them is of interest for the network. This helps in correct estimation of user speed giving important information about the consumption of network resources by user and his mobility. The cognition of speed will then help the network operator to perform online network resource and handover optimization according to users' speed profiles. It would be more precise if we predict the user speed itself. Nevertheless, from handover and resource optimization point of view, knowing the speed range instead of the speed itself can be sufficient. Thus, this paper proposes Mobile Speed Profiling (MSP) investigation. We mean MSP to be speed profiling by detecting speed range of an active user, in real time. MSP is done using real data by classifying the user speeds to multiple mobility states, during connection. For MSP, we show that we can achieve a good performance while collecting data with low and variable frequency. Whereas, for speed regression, this will require more data for training, with higher and constant frequency, which will increase the network overhead for collecting enough user-specific real data.

In literature, the issue of mobility state detection has been studied mainly considering 3 states: Low/ Medium/ High [2], [3]. However, mobile user speed profiling can require considering more than 3 states, when targeting to infer user mobility with more details. Actually, user's mobility conditions and thus speed deeply affects the way he interacts with his mobile phone. He usually tends to have different attitudes at home or at café than in transport, during the working week than during the week-end [4]. In [5] the authors observe this phenomenon inherent to the mobile user behavior: users prefer to use (or request service) their mobile phone in specific situations [6]. For more precision, we target a more detailed classification as compared to the one with only 3 mobility states.

In this paper, we propose MSP system using a supervised Deep Learning (DL) algorithm based on a multi-output classification. We detect multiple mobility speed profiles using multi-class schemes, i.e., up to N (we target to have $N \geq 3$) successive speed classes. Each class is bounded by a minimal and a maximal value. Questions to consider: to what detail such model requires to classify a user's speed? Can we classify the user's speed profile with detailed classes and with good performance? This paper will also analyse real data to show that the user's activity is correlated to the environment and speed categories selected for profiling. First, this enables us to highlight sets of classes with relevant labels for detailed MSP. Second, it allows us to provide a comparative analysis between various classification schemes involving multiple classes. The challenge is to find the class borders for the multi-output classification task for various degrees of granularity depending on the number of classes. The user speed needs to be profiled with good accuracy. However, one obstacle is that increasing the number of classes decreases the data instances in some classes and creates imbalanced data distribution. We will also show the trade-off between performance and more granularity.

Among AI approaches, DL is useful for multi-output classification. This is because it is appropriate for problems where modeling relationships between large number of features are not tractable. This is the case when classifying the mobility profiles into more than 3 states. Indeed the model has to extract the complexity and variety of different speed situations met by mobile users. DL is gaining much popularity due to its supremacy in terms of accuracy when trained with high amounts of data [7], [8]. The evaluation of the proposal will be done using real and large 4G LTE radio data collected through a crowdsourcing approach. Even if we are targeting 5G networks, the data collection has been done with existing materials, i.e. with 4G mobile phone device. Note that, this data is sent to the mobile network via 3GPP procedures compliant 5G network [9], making it applicable to 5G networks. Our paper's contributions are: 1) real data analysis for mobile speed profiling and also data cleaning algorithm 2) investigation of schemes for multiple classification of speed profiles 3) MSP detection based on DL.

The remaining paper is organized as follows. In Section II, past related works are discussed. Section III describes data collection and section IV investigates multiple schemes of speed profiling. In Section V a mobility speed profiling system is detailed and section VI discusses the results obtained when profiling the user speed. Finally, conclusions are presented.

II. Related work

In literature, many works have addressed the user mobility issue, mainly for handover management as well as for resource and optimisation reasons. Some works like [3] and [2] have studied the mobility speed profiling by determining the speed category of a mobile user. In [3], authors propose to use $RSRP$ (Reference Signal Receive Power) measurements to compute speed because signal fading over time is correlated with user speed. They propose 2 methods: one based on spectral analysis and other based on time-based spectrum spreading. In both methods, the signal variation is compared to a reference curve or a look-up table (database) and a mapping is used to compute the UE speed. Another method [2] for UE mobility estimation relies on UE history of cell sojourn times. The neighbouring eNBs exchange among them the learned network topology as well as the UE sojourn time history. Using such information, eNB classifies the speed to one of the three mobility classes defined by 3GPP. Such solutions are interesting and show good performances, but they are rather complex and resource-intensive (checking a reference database each time or exchanging information between eNBs).

Other works in literature have addressed the mobility issue, but they rather proposed solutions for estimating the future position or the future cell of a user. Moreover, they compute the speed using this information. In [10], authors use the knowledge extracted from simple features such as cell ID and sojourn time in previously visited cells, to predict the user mobility and his future location. The prediction of the future position of the mobile user is computed using SVM (Support Vector Machine). The work in [11] proposes a Markov chain based prediction technique to predict the next user position as well as his speed. Another work [12] uses a LSTM (Long short-term memory) based system to learn the user mobility pattern from historical trajectories and predict movement trends in the future. According to the prediction results of the next user position, both [12] and [11] propose an algorithm to optimize the handover management. According to Kipling's method, guessing the user pattern of next cell, does not really belong to the context awareness attributes. However, solutions proposed in these works, especially the ones using Machine Learning (ML) to detect mobility, are interesting in the context of our work. Another interesting work [13] evaluates the user mobility in 5G networks as a function of his behavior and his preferences. For evaluating mobility performance, they studied the user probabilities of pause, arrival, and departure.In this paper, we aim to study the user mobility from a context-awareness point of view. We focus on detecting user speed profile using a DL approach.

III. Data Collection

The collected data comes from a large crowdsourced campaign controlled by the operator, measured by phone devices and sent to the mobile network. The portable phone devices are always with users during their various activities with movement or no movement. This enables us to build a dataset for training and evaluation as close as possible to complexity and variety of a mobile user moving in real world. Thus, such collection mode allows to gather data inside the network corresponding to UE-specific radio measurements. It also includes the additional data required to do the labelling. Then this data is used to train the model in supervised mode.

As described in [14], our dataset has been collected at different locations (many cities and places) in France and on the roads linking these locations. Users are moving with different speeds. The data has been collected during 16 months, 24h/7, with an average of 1 measurement per 15 seconds, when the mobile phone session is active, and 1 measurement per 2 minutes otherwise. Around 2M lines of data per user have been collected. For MSP, a specific dataset is built for training. It consists of radio signals ($RSRP$, CQI (Channel Quality Indicator)), time related features (Timing Advance, Recording Time, Sojourn Time), Mobility Indicator (MI) and extra signals derived from the previous data. In addition to these signals described in [5], there is also other data. We used Global Positioning System (GPS) measurements for automatic labeling. The dataset is thus composed of above 10-features.

IV. Speed profiling schemes

In this section, we investigate the best speed categories to guarantee a better MSP to detect real user activity. For

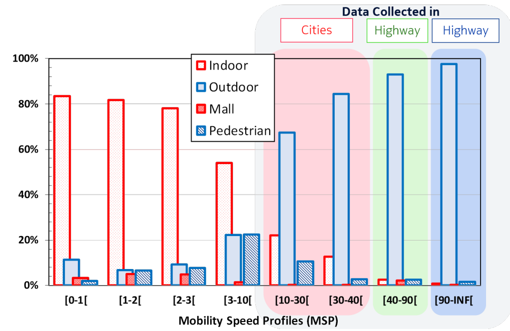

Fig. 1. Environment type distribution vs speed category (real data)

this, let us analyse user activity according to his or her speed profile and environment based on the real data collected. Later, we will study the MSP performance using different profiling schemes derived from the selected speed categories.

A. Speed category definition

Speed categories are highly influenced by many factors including the environment type which complicates the detection task. A speed category is defined by a minimal and maximal speed value. The boundaries of categories are extracted from the following set denoted as B :

- $B = \{0, 1, 2, 3, 10, 30, 40, 90\}$ (in kmph).

They were selected in order to reflect the complexity of a user's daily life and capture the variety of his movements in real world. They represent the typical speeds given in [15], corresponding to various environments met by mobile users (urban, rural, highways, road, pedestrian, bus, car and train).

The Fig. 1 draws the environment type distribution versus the speed category. Note that the borders shown are chosen from B. We observe that some boundaries are relatively much more visible between some speed categories according to the type of environment. We notice such prominent boundaries at 10 kmph and at 40 kmph. The 10 kmph boundary makes the split between indoor and outdoor that are two distinct environment types with different physical characteristics. The points below this boundary represent more than 60% (up to 80%) of measurement points collected inside of buildings. Whereas, the instances above 10 kmph have been mostly collected outdoor : in cities between 10 and 40 kmph and highway above 40 kmph. Note, there are errors as this is real data. This calls for data cleaning algorithms (proposed later).

B. Relation between user activity and speed profile

Fig. 2 depicts user activity by plotting the phone usage ratio for different speed categories derived from B. Note that a user's activity is characterized by his or her phone usage. The 'phone usage' corresponds to the state when the user uses his phone or equivalently the screen is on, as well as unlocked, and there is data exchanged. User activity per category is the ratio between the number of instances per category, when the user is using his phone, and the total number of activity instances. In fact, Fig. 2 illustrates the percentage of total time the user is connected to 4G network and is exchanging data.

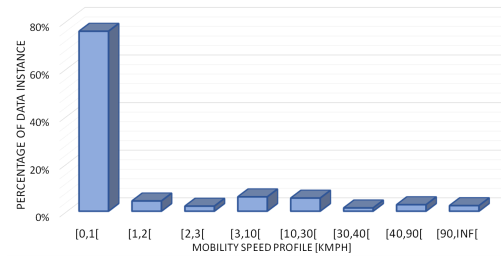

Fig. 2. User activity per mobility speed profile

We observe that most of the user activity occurs when the user moves at a speed lower than 1 kmph (75 % of activity), i.e, when he is indoor or is walking as a pedestrian.

Fig. 2 highlights the user activity trends that we observed after statistical analysis on mobile user behavior in literature. During day, a user is in different situations, such as walking outdoor, in a car, at work, in a mall, in transport or at home. Actually, mobile users' preferences for certain applications or contents are linked with his or her current usage situation [4]. In literature, some statistical studies show that mobile phones are mostly used when user is inside a building for internet service (80% of data calls) as well as for a call (70 % of voice calls) [6]. This can be explained by the fact that different use contexts pose their own limitations and impact the potential application usages. In [16], [17], the most commonly mentioned physical environments of application usage are indoor, but also include vehicles, such as public transportation and private cars. The home environment is preferred by users. Moreover, as soon as the user speed increases, the network quality deteriorates and consequently the user leaves 4G switching to 3G or even to 2G (rarely). Consequently, there exists more data in lowest speed category as compared to other categories, in terms of phone usage. We also observed in Fig. 2 that the data instances are distributed unequally between different categories. Thus, designing a speed classification scheme can face the problem of unequal classes in terms of data distribution. Moreover, today, this imbalanced nature of classes still remains a challenge for ML algorithms in general.

C. Classification schemes

The objective of this section is to define the relevant classification schemes by smartly regrouping various speed

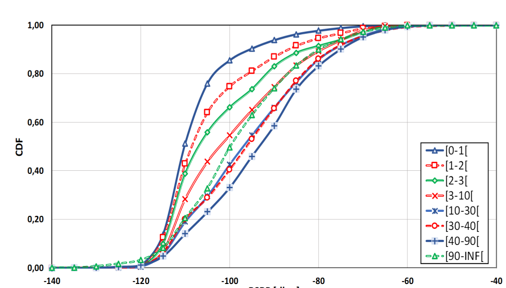

Fig. 3. Empirical CDF for measured RSRP per speed category

categories given in section IV-A. We aim to design a classification scheme that detects detailed speed profiles of mobile users, with a fine granularity and a low decision error margin. We assume this margin to be 5%. An investigation is needed to guess the relevant minimal value of this margin. But, it is out of scope of this paper.

We then consider different possibilities ranging from a simple 3-state classification problem to a more detailed classification of a user's speed. For that, we decide to regroup the speed categories based on the similarity between the cumulative distribution curves of the collected data $\{RSRP, MI\}$. These 2 features were identified to be contributing most after we ranked them using Principal Component Analysis method. The created merged groups shall also include data with similar statistical properties in order to optimize the classification results. Figures 3 and 4 depict the cumulative distribution curves of $RSRP$ and MI, respectively. We observed from these curves that 4 groups of similar curves can be extracted: a set with only $[0, 1[$ kmph, a set regrouping $\{[1, 2[$ and $[2, 3[\}$ kmph, a set alone with $[3, 10[$ kmph, and another assembling $\{[10, 30[$, $[30, 40[$, $[40, 90[$ and $[90, INF[\}$ kmph. This results into a clear split in 2 groups: one associated to users moving very slowly (walking or static) and the others moving at high speed. This separation is more noticeable in the CDF of MI.

Based on these observations, we propose to investigate ten schemes of 3, 4, 5 and 8 classes (3C, 4C, 5C and 8C). A class corresponds to a set of speed categories. Thus a class is defined by minimal and maximal boundaries of the associated speed categories. Let a multi-class scheme i be composed by a set of classes denoted as $\mathbb{C}^i = \{C_1^i, C_2^i, ...C_n^i...\}_{n \leq N_i}$ with N_i the number of classes. Let $Bound_i$ the set of boundaries coming from the speed categories selected. If, $Bound_i = \{B_1^i, B_2^i, ...B_j^i...\}_{j \leq N_i+1}$ with B_j the speed value in kmph. Then, the class n of the scheme i is written as: $C_n^i = [B_n^i, B_{n+1}^i[\ \forall\ n \leq N_i$ with C_n corresponding to speed values in the interval $[B_n, B_{n+1}[$.

A plurality of speed categories are then investigated to better model the diversity of speed situations, thus defined by the optimal boundaries of the categories:

- **3C**: $\{0, 10, 90\}$;$\{0, 40, 90\}$;$\{0, 3, 90\}$;$\{0, 3, 30\}$ kmph
- **4C**: $\{0, 1, 10, 90\}$;$\{0, 1, 3, 30\}$;$\{0, 1, 3, 90\}$ kmph
- **5C**: $\{0, 1, 2, 10, 90\}$;$\{0, 1, 10, 40, 90\}$ kmph
- **8C**: $\{0, 1, 2, 3, 10, 30, 40, 90\}$ kmph

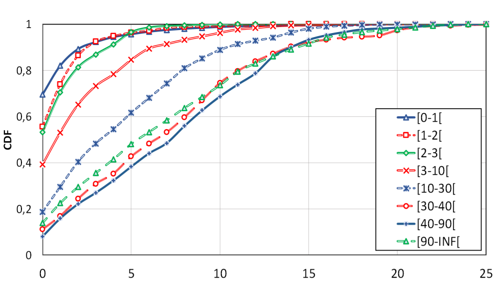

Fig. 4. Empirical CDF for measured MI per speed category

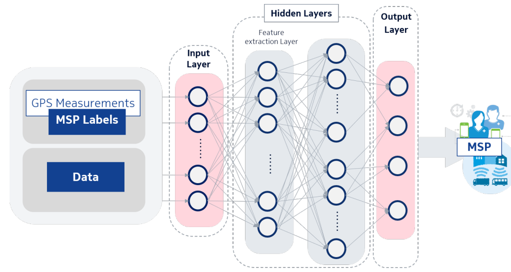

Fig. 5. Mobility Speed Profiling (MSP) system architecture - Note that there are several hidden layers and not just two as it may initially appear.

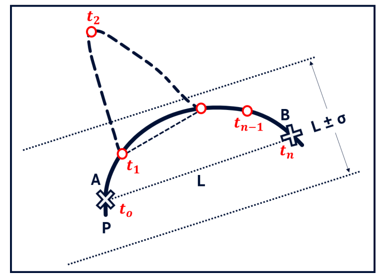

Fig. 6. Speed Computing between according to path P between two points A and B during a sliding window TCR_{max}

V. Mobility Speed Profiling System

A. System architecture

In reality, human behavior is complex to model mathematically because it is unpredictable and doesn't follow a well-defined logic. DL is a sub-field of ML where we parse historical data to teach the machine to solve a complex task that may be difficult to model mathematically.

We propose to investigate DL-based model for MSP. The system architecture is shown in Fig. 5. It is a Feed Forward Neuronal Network (FNN) composed of 3 main parts :

- Input: A first input layer is fed with 10-features described in section III that are 4G collected data.
- Core: Containing 5 hidden layers as well as a dropout layer to regularize and minimize the over-fitting.
- Output: An output layer with either 3, 4, 5 or 8 classes.

B. Data cleaning method for labeling

Actually, GPS measurements are often used in the context-aware characterization. Indeed, such measurements give a good information about the user location at a time t. Thus, a typical approach is to transform the series of measurements that record position points (latitude and longitude) at regular time intervals in coordinates (x, y) in km. To derive the average mobile user speed v, we use a succession of N coordinates (x, y) derived from GPS information. Note that this is only done during data labeling phase and GPS information is not assumed to be available during classification phase. Imagine that the mobile user moves along a path P from point A to point B (Fig. 6). To link these two extreme points the user goes through $N - 2$ intermediate points belonging to $S = \{a_i\}_{i \leq N-1}$ at time $\{t_2\}_{i \leq N-1}$.

Let TCR_{max} be the elapsed time to go from A to B and L the total distance. Let δt_i the elapsed time and l_i the distance between $\{a_{i-1}, a_i\}$. The point a_i has the following coordinates $(x(t_i), y(t_i))$. $l_i = \sqrt{(x(t_i) - x(t_{i-1}))^2 + (y(t_i) - y(t_{i-1}))^2}$ and $\delta t_i = t_i - t_{i-1}$. The average speed $\hat{v}$, is approximated as:

$$\hat{v} = \frac{1}{N-1} \sum_{i=1}^{N} \frac{l_i}{\delta t_i} = \frac{1}{N-1} \sum_{i=1}^{N} \hat{v}_i$$
$$\text{with } t_N - t_0 = TCR_{max}$$

As we use real data, the GPS measurements are noisy, namely some data are erroneous. For this reason, GPS data should be cleaned in order to ignore measurements that we detect as erroneous. This step is primordial since it ensures the integrity of the ground truth used for labeling. To do this, let σ be the threshold that sets a confidence interval. At t_i, if $l_i \in [L \pm \sigma]$ then $a_i \in$ S otherwise $a_i \notin$ S. The distance l_i between 2 successive points should be bounded by L to be used for the average speed derivation. Otherwise the measurement at time t_i is considered as outlier and is excluded. The label computing is detailed in Algo. 1.

To compute the speed, the question is: what is the most appropriate value of TCR_{max} (width of the sliding window) to compute the speed? Considering micro-cells, mostly represented in urban environments, we assume typical cell radius distances between 0.2km and 2km. For both values, curves of mobile user speed versus cell crossing time is plotted in Fig. 7. They enable us to fix TCR_{max} at 10 kmph where there is a clear separation between CDFs of MI. So, the duration TCR_{max} for calculating the speed is empirically set to $300s$.

Algorithm 1 LABEL COMPUTING

$(x, y) \leftarrow$ transform(lat,long,TCR_{max})
$L_{AB} \leftarrow \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$
Initialize σ, $Index \leftarrow 0$ And $v \leftarrow 0$
for i in $[0, TCR_{max} - 1])$ **do**
    $l_i \leftarrow \sqrt{(x_{i+1} - x_{Index})^2 + (y_{i+1} - y_{Index})^2}$
    if l_i in $[L_{AB} \pm \sigma]$ **then**
        $v \leftarrow v + \frac{l_i}{t_{i+1} - t_{Index}}$ And $Index \leftarrow Index + 1$
    else
        Consider Measure i as an outlier and skip it
    end if
end for
$V_{TCR_{max}} \longleftarrow \frac{v}{Index+1}$
Label Coding according to $v_{TCR_{max}}$

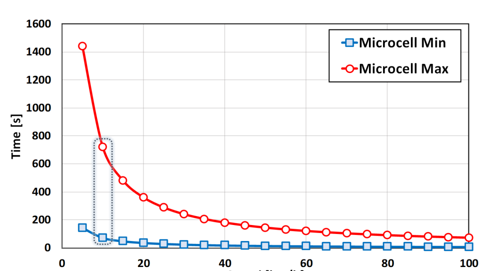

Fig. 7. Speed versus time for two cell radius

VI. EXPERIMENTAL SETUP, RESULTS AND DISCUSSION

A. Configuration and performance metrics

For our supervised multi-output classification problem, we used a FNN architecture with a total of 5 hidden layers that is depicted in Fig. 5. The MSP is implemented with both Scikitlearn and Keras using python with tensorflow as the back-end. A crucial step of DL is to optimize the hyper-parameters set for the model. The set of hyper-parameters are e.g. the number of hidden layers, batch size, epoch size, the weights, the activation function, the loss function, the learning rate etc. Recently, a new approach using a Bayesian optimization for tuning hyper-parameters has been considered [18]. The Bayesian method has the property to rapidly reach the optimal set of hyper-parameters than others, thus we choose this method to tuned those of our model.

The FNN model is built with around 300k instances of LTE data collected only in France. It has been trained using 70% and tested on the remaining 30%. As the MSP problem leads to an issue of imbalanced classes, F1-score metric, in addition to the accuracy, is also used for evaluation. The metric by definition is the weighted average of *Precision* and *Recall* according to the following relation: $F1 - score = \frac{2.Precision.Recall}{Precision+Recall}$

B. Results

First, in order to quantify the added value of correcting labels, we trained the model with and without correcting labels on the scheme $\{0, 40, 90\}$ as it was studied in [2]. Results show a gain of around 15% (From 78.3% to 93.8% of F1-score) while using the label correction Algo. 1. Indeed, the performance increases with better data and labelling.

Table VI presents the performance results of the multi-output classification schemes. We observe that we have an overall good performance. All the classification schemes considered show an accuracy higher than 93% and an average F1-score of 90%. However, we also see that the higher the number of classes is, the lower the performance is. This can be explained as follows: increasing the number of classes increases the complexity of the model as well as the relations between the features and the output. In addition, higher number of classes reduces the number of instances of data in each class which leads to a problem of imbalanced classes. Indeed, imbalanced classes can pose a problem during the training phase of any ML algorithm and not only for neural networks [19]. The effect of imbalanced classes increases if the boundaries between classes are very similar, which is the case for us. To resolve such issue, we employed another technique called Artificial Data Augmentation (A-DA). A-DA learns the distribution of a stratified sample from our dataset and then generates data with similar statiscal properties. However, we limited the generation of the artificial data to 65k in order to keep close to the original data and to avoid additional noise. The impact of artificial data augmentation (A-DA) on the best 3 classification schemes with FNN (cases in bold) shows an improvement of around 3.0% in terms of F1-score. This is explained by the fact that FNN is greedy: more data results in better performance [7], [8].

	3 Classes								4 Classes	
	$[0, 10, 90]$		$[0, 40, 90]$		$[0, 3, 90]$		$[0, 3, 30]$		$[0, 1, 10, 90]$	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
FNN	**98.367%**	**94.256%**	98.222%	93.861%	97.226%	93.570%	96.956%	93.038%	**95.365%**	**91.167%**
FNN + A-DA	**98.096%**	**96.012%**							**95.968%**	**93.559%**
	4 Classes				5 Classes				8 Classes	
	$[0, 1, 3, 30]$		$[0, 1, 3, 90]$		$[0, 1, 2, 10, 90]$		$[0, 1, 10, 40, 90]$		$[0, 1, 2, 3, 10, 30, 40, 90]$	
	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
FNN	95.004%	89.917%	95.517%	90.447%	94.484%	87.285%	**94.913%**	**89.066%**	**93.321%**	**81.108%**
FNN + A-DA							**94.283%**	**92.092%**	**93.933%**	**85.484%**

TABLE I

DEEP LEARNING-BASED SUPERVISED MULTI-OUTPUT CLASSIFICATION PERFORMANCE USING A FEED FORWARD NEURONAL NETWORK (FNN) WITH AND WITHOUT ARTIFICIAL DATA AUGMENTATION (A-DA): F1-SCORE VS. CLASSIFICATION SCHEMES

We observe that certain common boundaries are found in schemes with higher MSP performance. The best schemes per class (in bold) are those which fix the boundaries to 1, 10, 40 and $90kmph$. Indeed, $10kmph$ makes the split between indoor and outdoor. Therefore classes on both sides of $10kmph$ are well detected because they are differentiated by the environment of the location where data was collected. Furthermore, $[10, 30[$, $[30, 90[$ and $[90, INF[$ present various high speed profile environments (cities and highways in urban as well sub-urban). Moreover, introducing $1kmph$ boundary brings enhancement. Indeed, instances in the $[0, 1[$ class are very well detected since this class is the most populated class. But, the smallest classes $[1, 2[$, $[2, 3[$, $[3, 10[$ are detected with lesser accuracy since they mostly present similar physical properties and are confused with $[0, 1[$. Thus, we show using real data that a detailed learning of the user mobility speed profiles can be achieved with good performance.

VII. CONCLUSION

To deliver context-aware services, it is important to first detect the use context. In this paper, we have investigated the mobility speed profiling (MSP) of the user while consuming a service. We have proposed an intelligent system for MSP detection linked with user activity preference. We used a FNN network to learn the user mobility profiles when he is active and using his phone. Based on crowdsourced user-specific data, we achieved above 94.5% of detection in terms of F1-score in existing deployed 4G HetNets. We worked with real data using GPS coordinates as labels. They must be cleaned because of inherent noise. This noise can impact the overall performance of any ML algorithm. We observed that the speed profiling is influenced by the user environment. This is because signals used for MSP are impacted by the environment of the data collection location. Thus, for learning more details, it is necessary to pay attention to the environment type in order to fix the boundaries for the MSP task. Furthermore, we noted that the level of imbalance in data influences the F1-score. One way to overcome this issue is to collect more data or slightly re-balance the data artificially. In future, we will evaluate the performance of our intelligent MSP system on more data and more diversified profiles. We will also extend and optimize our work by introducing the temporal as well as the spatial correlation between the features using other DL architecture as CNN (Convolutional Neural Network) and LSTM.

REFERENCES

[1] JAVAID Nadee, SHER Arshad, NASIR Hina Nasir, GUIZANI Nadra. Intelligence in IoT-Based 5G Networks: Opportunities and Challenges. In : IEEE Communications Magazine, October 2018. p. 94-100.

[2] HERCULEA, Dalia, CHEN, Chung Shue, HADDAD, Majed, et al. Straight: Stochastic geometry and user history based mobility estimation. In : Proceedings of the 8th ACM International Workshop on Hot Topics in Planet-scale mObile computing and online Social neTworking. ACM, 2016. p. 1-6.

[3] HADDAD, Majed, HERCULEA, Dalia Georgiana, ALTMAN, Eitan, et al. Mobility state estimation in LTE. In : Wireless Communications and Networking Conference (WCNC), 2016 IEEE. IEEE, 2016. p. 1-6.

[4] S. Ickin et al., Factors influencing quality of experience of commonly used mobile applications, IEEE Communications Magazine, vol. 50, no. 4, pp. 48-56, 2012.

[5] Marie-Line Alberi-Morel, Illyyne Saffar, Kamal Singh et al. Multi-task Deep Learning-based Environment and Mobility Detection for User Behavior Modeling. International Workshop on Machine Learning for Communications (WMLC 2019) France.

[6] http://www.hetnetforum.com/, HetNet Forum, 2018

[7] ZHANG, Chaoyun, PATRAS, Paul, et HADDADI, Hamed. Deep Learning in Mobile and Wireless Networking: A Survey. arXiv preprint arXiv:1803.04311, 2018.

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

[9] GPP TS 3GPP TS 38.305: "NG Radio Access Network (NG-RAN); Stage 2 functional specification of User Equipment (UE) positioning in NG-RAN", Release 15

[10] MICHAELIS, Stefan. Balancing high-load scenarios with next cell predictions and mobility pattern recognition. Dortmund, 2012.

[11] ULVAN, Ardian, ULVAN, Melvi, et al. The enhancement of handover strategy by mobility prediction in broadband wireless access. In : Proceedings of the networking and electronic commerce research conference (NAEC 2009).

[12] WANG, Chujie, ZHAO, Zhifeng, SUN, Qi, et al. Deep Learning-based Intelligent Dual Connectivity for Mobility Management in Dense Network. In : 2018 IEEE 88th Vehicular Technology Conference 2019.

[13] GE,Xiaohu, YE,Junliang, YANG, Yang, et al. User mobility evaluation for 5G small cell networks based on individual mobility model. IEEE Journal on Selected Areas in Communications, 2016.

[14] SAFFAR, Illyyne, ALBERI-MOREL, Marie-Line, SINGH, Kamal Deep, et al. Semi-supervised Deep Learning-based Methods for Indoor Outdoor Detection. IEEE International Conference on Communications.

[15] 3GPP TS 3GPP TS 38.913: "Study on scenarios and requirements for next generation access technologies", Release 14

[16] V. A. Siris, K. Balampekos and M. K. Marina, Mobile Quality of Experience: Recent Advances and Challenges, Workshop on Information Quality and Quality of Service for Pervasive Computing, 2014

[17] T. Soikkeli, et al. Diversity and End User Context in Smartphone Usage Sessions, Next Generation Mobile Applications, Services and Technologies (NGMAST), 5th International Conference, 2011.

[18] BERGSTRA, James, YAMINS, Dan, et al. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms.

[19] GÓMEZ, Santiago Egea, HERNÁNDEZ-CALLEJO, et al. Exploratory study on Class Imbalance and solutions for Network Traffic Classification. Neurocomputing, 2019.