Neurocomputing 358 (2019) 479–489

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Brief papers

A novel deep learning driven, low-cost mobility prediction approach for 5G cellular networks: The case of the Control/Data Separation Architecture (CDSA)

Metin Ozturk[a,*], Mandar Gogate[b], Oluwakayode Onireti[a], Ahsan Adeel[b,c], Amir Hussain[b], Muhammad A. Imran[a]

[a] School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK
[b] School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, UK
[c] School of Mathematics and Computer Science, University of Wolverhampton, Edinburgh EH16 5XW, UK

ARTICLE INFO

Article history:
Received 11 June 2018
Revised 22 October 2018
Accepted 7 January 2019
Available online 17 January 2019

Keywords:
Mobility management
Deep learning
Predictive handover
5G wireless mobile networks
Control/Data Separation Architecture

ABSTRACT

One of the fundamental goals of mobile networks is to enable uninterrupted access to wireless services without compromising the expected quality of service (QoS). This paper reports a number of significant contributions. First, a novel analytical model is proposed for holistic handover (HO) cost evaluation, that integrates signaling overhead, latency, call dropping, and radio resource wastage. The developed mathematical model is applicable to several cellular architectures, but the focus here is on the Control/Data Separation Architecture (CDSA). Second, data-driven HO prediction is proposed and evaluated as part of the holistic cost, for the first time, through novel application of a recurrent deep learning architecture, specifically, a stacked long-short-term memory (LSTM) model. Finally, simulation results and preliminary analysis reveal different cases where non-predictive and predictive deep neural networks can be effectively utilized, based on HO management requirements. Both analytical and machine learning models are evaluated with a benchmark, real-world dataset measuring human behaviors and interactions. Numerical and comparative simulation results demonstrate the potential of our proposed deep learning-driven HO management framework, as a future benchmark for the mobile networking and machine learning communities.

1. Introduction

Next generation wireless cellular networks are envisioned to be self-organized, efficient, and cost-effective [1]. Since 5G Self-Organizing Network (SON) is a new paradigm, there exist numerous design challenges, ranging from mobility management to resource management for seamless access to wireless services without compromising the expected Quality of Service (QoS).

Mobility management has two main subsets: handover (HO) management and location management. The former covers aspects of users Access Point (AP) or Base Station (BS) switching, while the latter is based on location tracking of a user. HO prediction is one of the most widely used approaches in HO management for cellular networks since it allows proactive radio resource allocation. Accurate HO prediction can significantly reduce HO latency, signaling overhead, and call drop rate.

The Control/Data Separation Architecture (CDSA) is a promising, enhanced network architecture, with a logical separation between the Control Plane (CP) and Data Plane (DP). The CDSA assigns CP for coverage provisioning, while DP is responsible for data transmission [2]. A typical CDSA with one Control Base Station (CBS) and four Data Base Stations (DBSs) is shown in Fig. 1.

CDSA offers a number to advantages over conventional cellular architectures including better energy efficiency and system capacity. In addition, it prevents DBSs from excessive overheads [2]. This is mainly because CP and DP are combined in the conventional architecture, where each BS is responsible for coverage, mobility, and data transmission tasks, that leads to a signalling overhead. However, in CDSA, CBSs are responsible for connectivity and mobility issues, thus preventing DBSs from excessive overhead.

In this paper, a novel analytical model for holistic HO cost evaluation, taking into account signaling overhead, latency, call dropping, and radio resource wastage, is presented. In addition, novel stacked long-short-term-memory (LSTM) and multi-layered perceptron (MLP) based mobility prediction models are proposed

* Corresponding author
E-mail address: m.ozturk.1@research.gla.ac.uk (M. Ozturk).

https://doi.org/10.1016/j.neucom.2019.01.031

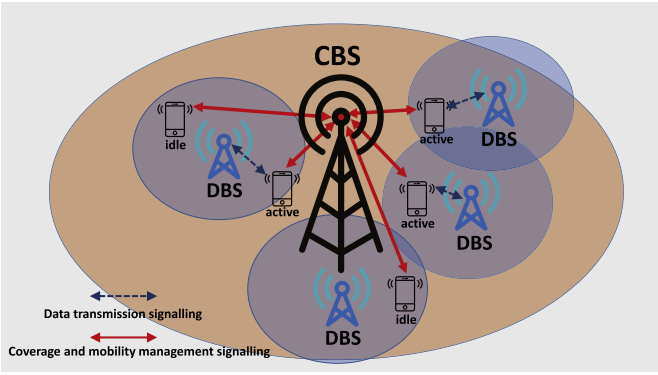

Fig. 1. Sample CDSA architecture. There are four DBSs under the coverage of one CBS. User Equipments (UEs) can be in active and idle states according to their DBS connection.

and evaluated as a part of the holistic cost function. Comparative simulation results depict the superiority of LSTM over MLP, for accurate HO prediction. The performance improvement can be attributed to LSTM's inherent recurrent nature, that helps better model time series data.

In summary, this paper reports three major contributions outlined below:

1. Development of a novel analytical model for holistic HO cost evaluation model, by integrating signaling overhead, latency, call dropping, and radio resource wastage. The developed mathematical model is applicable to several cellular architectures, including CDSA.
2. A HO prediction algorithm is proposed to further minimize the holistic cost. It is shown that high prediction accuracy can significantly minimize the cost function, consequently reducing user dissatisfaction, HO latency, overhead, and resource wastage.
3. The impact of prediction accuracy as part of the holistic cost is evaluated to estimate HO accuracy requirements, considering the holistic cost, and a novel deep neural network based learning model is developed for predictive HO management. It is shown that the predictive approach significantly outperforms the non-predictive method, for the case where the former offers more than the required minimum accuracy. On the other hand, when the achieved prediction accuracy equals the required minimum accuracy, both predictive and non-predictive approaches perform equally well. However, non-predictive approaches outperform predictive methods when the required minimum accuracy is not achieved by the machine learning algorithm. Thus, a future context-aware approach is posited, in which the mobile network can optimally switch between predictive and non-predictive approaches, based on HO management requirements.

The remainder of this paper is organized as follows: Section 2 provides an up-to-date literature survey, while Section 3 presents the background information on HO elements. The holistic HO cost and minimum requirements for prediction accuracy are derived in Section 4. Section 5 presents data-driven HO prediction models based on shallow MLP and deep stacked LSTM networks. Comparative performance evaluation and results are discussed in Section 6. Lastly, Section 7 concludes the paper and outlines some future work directions.

2. Related work

In the literature, machine learning has been widely used to address the issue of HO prediction in mobile networks. Specifically, Markov Chains [3–8] and Artificial Neural Networks (ANN) [9–11] have been extensively employed. More specifically, in [3], authors employed two different methods to predict mobile users' trajectories. Specifically, Dempster-Shafer processes were used to estimate the next location of users, while a second-order Markov chain was employed at road segments to determine the segment the user would take.

Online Markov chains were proposed in [4,5] to decrease the signalling overhead and latency caused by HOs. Given the serving BS of a mobile user, the algorithms predict the target BS by executing the proposed online learning algorithm. The revisit problem of Markov chains, when applied to mobile user's trajectory prediction, was identified in [6]. As a solution for this problem, the authors proposed to build a 3D transition matrix instead of the conventional 2D, in order to distinguish the orders of HOs. The authors in [7] model the mobility behaviour of users with Markov Renewal Process (MRP) to predict both single and multi-transitions. The case of having close probabilities in the transition matrix, which can cause serious accuracy degradation, were studied in [8], and the concept of not making a prediction in case of not having *enough* assurance was introduced. The term *enough* is represented by a threshold value, and if the probabilities of the candidate states cannot meet the threshold value, the agent chooses not to make a prediction.

ANNs are employed in [9,11] to learn user patterns and use them for future location predictions in UMTS networks, thereby decreasing the location update signalling overhead. The Self-Organizing Map (SOM) algorithm is utilized in [10] for femtocells to learn and decide if HOs are taking place inside the indoor environment necessary. Thus, the ping-pong HOs taking place among femto and macro cells can be avoided.

However, none of these studies have explored a holistic cost function for HO management, although individual cost has been considered in a few studies. Lack of this global cost assessment is thus an open challenge, with no complete evaluation available, other than a number of individual key performance indicators (KPIs).

Recently, several heuristic approaches have also been employed. For example, authors in [12] proposed a hybrid intelligent HO decision algorithm based on Artificial Bee Colony and Particle Swarm Optimization, in order to select an optimal network during vertical HOs. Similarly, a hybrid Markov-based model for human mobility prediction is presented in [13]. Nevertheless, real-time communication requires algorithms that are fast and effective in selecting optimal available networks. Therefore, low computational complexity, real-time learning, and optimization algorithms are required, such as those recently proposed in [14,15]. One of the main challenges of designing a real-time cognitive system is to concurrently acquire long-term learning, fast decision making, and low computational complexity. The authors in [16] developed a novel random neural network based optimization system. However, such real-time AI driven optimization engines are yet to be practically exploited.

Deep learning implementations in mobile networks are not limited to these aforementioned studies. In [17], for example, the authors extensively studied a deep learning formulation for the mobile encrypted traffic classification problem. Various deep learning architectures, presented in the literature, are reproduced, trained, and tested with three different data sets, and a comprehensive performance comparison in terms of accuracy, precision, and recall is provided.

In general, the deep learning framework has a broad application area, with a growing number of state-of-the-art works reported to-date. For example, recent work in [18] exploited Deep Sparse Autoencoders (DSAE) to recognize facial expressions. The proposed method was tested with a data set and attained 95.79% accuracy.

In [19], Deep Belief Networks (DBF) was applied to produce more qualitative analysis of gold immunochromatographic strip (GICS) images. The objective was to differentiate the test and control lines in GICS images, and very high accuracy levels were achieved for various concentrations. Both [18,19] are representative case studies depicting the capability of deep learning algorithms as universal approximators.

3. Background

HO management has been one of the main challenges for cellular wireless networks. Cellular networks comprise cells where each cell covers a certain area with a BS. Thus, when a user traverses from one cell to another, all related information should also be transferred to the new BS, with an effective management strategy and coordination. However, this coordination causes delays and signalling overhead due to required signaling between BSs and UE, and BSs and the core network. Moreover, in dense small cells deployment, HO management becomes more challenging due to the large number of required HOs.

Predictive HO management is one possible solution to address this issue, by enabling proactive HO, due to known future locations of users. This, in turn, can mitigate the latency and signalling overhead significantly.

However, proactive HO management suffers from inefficient radio utilization, as it reserves resources for upcoming HO users. In dense network deployments, resource reservation becomes even more problematic since radio resources are already scarce in such scenarios. Furthermore, wrong predictions result in adverse consequences, as reserved resource will not be utilized. On the other hand, advance resource reservation with predictive HO management, decays the call dropping probability [20–22]. Thus, there are a number of elements to consider before employing a predictive HO process, such as latency, signalling overhead, call dropping probability, and radio resource wastage.

3.1. HO latency

HO latency is defined as time spent during the whole HO process, which includes preparation, execution, and completion phases. As shown in Fig. 2, this process for the conventional, non-predictive method starts with the measurement report and then follows steps from 3 to 12. Each of these messages among nodes (UE, DBS, MME, S-GW) takes some time to execute, which leads to HO latency.

3.2. Signalling overhead

The signalling overhead also results from messages among aforementioned nodes during the HO. Therefore, there is a direct relationship between latency and signalling overhead, since both result from messaging during the HO process. In other words, increasing the latency increases the signalling overhead, and vice versa. Moreover, the authors in [4,5] expressed the signaling overheads in terms of HO latency.

3.3. Predictive HO and call dropping ratio

The predictive HO concept offers in-advance preparation for upcoming HOs by employing machine learning or data mining tools. On the other hand, this in-advance preparation requires reserving the radio resources for HO users before HOs take place. Besides, these reserved resources are not allowed to be used by any other user. Hence, reserved resources are useless until HOs occur, which is termed radio resource wastage [4]. Although advance resource reservation is inconvenient in terms of resource wastage, it is quite utilitarian for the call dropping ratio.

The probability of call dropping depends on a variety of reasons, such as network density, data traffic, user profile, etc. The urban scenario, for example, has higher call dropping probability than the rural scenario, as more intense network density, which causes more data traffic, leads to more scarce resources.

In this regard, the predictive HO process offers a potential solution to decrease call dropping probability in the correct prediction case [20–22], by reserving resources for HO users in advance. Even if there is lack of resources at the predicted DBS, for a HO user by the time the prediction is made, it is still possible to find enough resources before HOs are performed. In other words, the correct prediction case permits DBSs to arrange the required resources before HOs occur. Thus, as the DBS has time (from prediction to HO trigger) to be prepared, there is a lower probability of not having enough resources for HO users.

Consequently, the benefits of predictive HO should be evaluated by including all these elements jointly. In the next section, we provide our system model for a holistic HO cost, and minimum, prediction accuracy requirement evaluations.

4. System model

In this study, joint cost models for both predictive and non-predictive HO procedures are derived. Furthermore, the aforementioned individual cost elements are distinctive for different scenarios, for example, user dissatisfaction could be a priority for rural areas, while resource wastage cannot be tolerated for ultra-dense networks. There are two variations of predictions in the predictive HO management: correct and incorrect predictions; hence, there will be three different cost models namely, for the non-predictive case, correct and incorrect prediction cases.

We already know from Mohamed et al. [4,5] that, on one hand, predictive HO with correct prediction can alleviate the signalling overhead as well as the HO latency, while on the other hand, it inflates the radio resource wastage. Moreover, from Zhang et al. [20–22], the call dropping ratio can be mitigated with an accurate predictive HO process. Overall, the holistic HO cost should include all these elements as follows:

$$C_{\mathrm{HO}} = pC_{\mathrm{D}} + (1-p)[C_{\mathrm{L}} + C_{\mathrm{OH}} + C_{\mathrm{W}}], \tag{1}$$

where p is the probability of not having enough resources at the target DBS, which will turn out to be a call drop. C_{D}, C_{L}, C_{OH}, and C_{W} are the costs for call dropping, latency, signalling overhead, and resource wastage, respectively.

All aforementioned scenarios; e.g. no-prediction, correct and incorrect predictions, have their own characteristics for the elements in (1). In the following sections, we investigate each parameter individually for all scenarios.

4.1. HO latency

The concept of HO latency, which is defined as the time required for the whole HO process [23], can be written as follows:

$$L_{\mathrm{HO}} = \begin{cases} t_{\mathrm{p}} + t_{\mathrm{e}} + t_{\mathrm{c}} + 100\mathrm{ms}, & if \text{ cell unknown} \\ t_{\mathrm{p}} + t_{\mathrm{e}} + t_{\mathrm{c}} + 20\mathrm{ms}, & \text{otherwise} \end{cases} \tag{2}$$

where t_{p}, t_{e}, and t_{c} are time required in the without-prediction case for HO preparation, execution, and completion, respectively. If the target cell is unknown, a search delay, set to 80 ms by 3GPP [24], is added to the budget in addition to a 20 ms margin. Moreover, since the HO delay is being investigated from the UE's point of view, t_{c} should be counted as zero, as its Radio Resource Control (RRC) connection is performed when the HO completion phase starts [23].

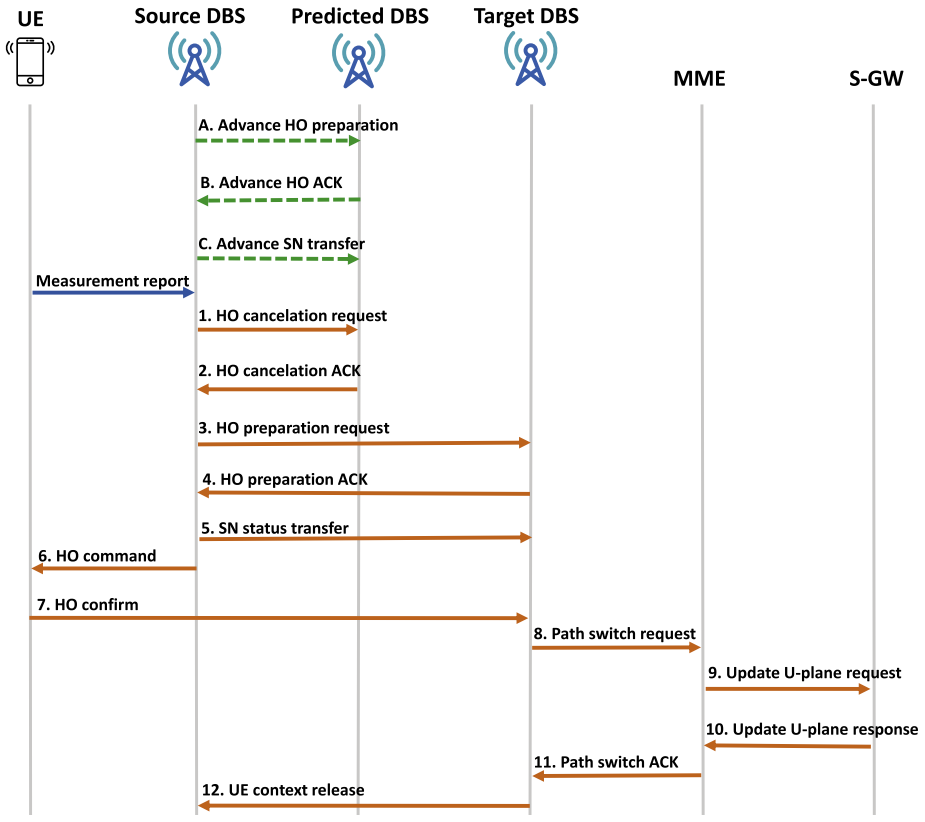

Fig. 2. LTE X2 HO procedure as in [4]. Non-predictive HO conducts steps from 3 to 12. A predictive process with correct prediction performs steps from 6 to 12, whereas steps from 1 to 12 are taken in the case of an incorrect prediction. Steps *A*, *B*, and *C* are for the correct prediction case, and are performed in advance.

As we can reduce the latency of the preparation phase only [5], therefore, it can be assumed as a latency cost of HO:

$$C_{\mathrm{L}} = \begin{cases} t_{\mathrm{p}} + 100\mathrm{ms}, & if \text{ cell unknown} \\ t_{\mathrm{p}} + 20\mathrm{ms}, & \text{otherwise} \end{cases} \tag{3}$$

With the help of correct predictions, t_{p} can be reduced; such that if, for example, the future locations of a user are known, preparations for upcoming HOs can be done in advance by skipping some of the steps performed in the conventional HO. It is noted in [5] that a predictive HO with correct prediction is better than the conventional process in terms of the number of steps taken during the HO process. However, whenever an incorrect prediction occurs, due to resources being allocated to the wrong cell, the conventional process is better than the predictive one. This implies that making an incorrect prediction incurs a penalty in terms of HO latency.

Overall, t_{p} can be expressed as follows for three different scenarios [5]:

$$C_{\mathrm{L}} = \begin{cases} C_{\mathrm{L,NP}}, & for \text{ no prediction(NP)} \\ C_{\mathrm{L,IP}}, & for \text{ incorrect prediction (IP)} \\ C_{\mathrm{L,CP}}, & for \text{ correct prediction(CP)} \end{cases} \tag{4}$$

where $C_{\mathrm{L,CP}} < C_{\mathrm{L,NP}} < C_{\mathrm{L,IP}}$. If the LTE X2 HO procedure is used and values in [5] are adopted, the corresponding costs become as follows: $C_{\mathrm{L,CP}} = 20$ ms, $C_{\mathrm{L,NP}} = 35$ ms, and $C_{\mathrm{L,IP}} = 45$ ms. Note that we assume the target cell is known, thus t_{search} in [24] is zero.

4.2. Signalling cost

The HO signalling cost is a combination of transmissions costs caused by the messaging between DBSs, DBS and UE, and DBS and Mobility Management Entity (MME); processing costs at the DBS, MME, and Serving Gateaway (S-GW); and UE's detaching and access costs. Therefore, it is worth noting that the signalling cost mentioned here is from both UE's and the network's point of view [5]. The signalling cost is defined as [25]:

$$C_{\mathrm{OH}} = C_{\mathrm{s}} + C_{\mathrm{m}}, \tag{5}$$

where C_{s} and C_{m} are the signalling costs for search and movement, respectively. Similar to the latency, the signalling cost occurs at every phase of the HO process, thus it is again safe to consider only the preparation phase, that is the only phase is subject to a reduction.

$$C_{\mathrm{OH}} = \begin{cases} C_{\mathrm{OH,NP}}, & for \text{ NP} \\ C_{\mathrm{OH,IP}}, & for \text{ IP} \\ C_{\mathrm{OH,CP}}, & for \text{ CP} \end{cases} \tag{6}$$

In this work, we propose a conceptual signalling cost model that reflects the number of messages to be transferred during the process. To realize this, we assume each messaging among the same bodies causes the same signalling cost. For example, the signalling cost for messaging between the source DBS and the target DBS, and between the source DBS and the predicted DBS are assumed to be the same, since both take place among DBSs. Therefore, the relationship between all three cases in terms of the signalling cost is $C_{\mathrm{OH,CP}} < C_{\mathrm{OH,NP}} < C_{\mathrm{OH,IP}}$, as shown in Fig. 2. Since steps 1 to 5 constitute the HO preparation phase (messaging between DBSs), the number of messages that represent the signalling costs are as follows: $C_{\mathrm{OH,CP}} = 0$, $C_{\mathrm{OH,NP}} = 3$, and $C_{\mathrm{OH,IP}} = 5$.

4.3. Resource wastage

The resource reservation is performed after triggering HOs in conventional non-predictive HO management. In contrast, in

predictive HO, it is performed before the HOs are triggered. Moreover, in the case of an incorrect prediction, the dimension of the wastage is larger, since reserved resources in wrongly predicted DBS are never used. In such cases, non-predictive HO is better than the predictive one, even if it is a correct prediction scenario. The resource wastage cost is defined as follows, for no-prediction, incorrect prediction, and correct prediction cases, respectively:

$$C_{\mathrm{W}} = \begin{cases} C_{\mathrm{W,NP}}, & for\ \mathrm{NP} \\ C_{\mathrm{W,IP}}, & for\ \mathrm{IP} \\ C_{\mathrm{W,CP}}, & for\ \mathrm{CP} \end{cases} \tag{7}$$

where $C_{\mathrm{W,NP}} < C_{\mathrm{W,CP}} < C_{\mathrm{W,IP}}$.

In this work, resource wastage is modelled as the product of the demanded resource and reservation time. If the resource wastage of the non-predictive process is assumed to be zero, the resource wastage model can be given as follows: $C_{\mathrm{W,NP}} = 0$, $C_{\mathrm{W,CP}} = t_{\mathrm{a}} d_{\mathrm{r}}$, and $C_{\mathrm{W,IP}} = t_{\mathrm{m}} d_{\mathrm{r}}$, where d_{r} is a demanded resource, t_{a} is time elapsed between the prediction and the HO trigger, and t_{m} is the maximum possible time that the resources are reserved for the predicted HO. In other words, the former is the time for the correct prediction case, while the latter is exclusive to the incorrect prediction scenario. Hence, the t_{m} parameter ensures the predicted DBS reserves demanded resources for a certain time, and then releases them if the planned HO does not happen. Eventually, it is worth noting that $t_{\mathrm{m}} > t_{\mathrm{a}}$.

4.4. Probability of call dropping

When a HO is not managed properly, it ends up dropping a call, causing user dissatisfaction. In other words, if the target DBS has insufficient resources to accommodate the HO user, then call must be dropped. In addition, according to [26], call dropping is less tolerable than call blocking, which happens when the network fails to accommodate new calls due to being short of resources.

However, incorrect prediction does not have any impact on the dropping probability, as the network conducts the conventional HO process in the target DBS when the prediction is incorrect. Therefore, the call dropping probabilities for non-prediction and incorrect prediction cases are the same, while the correct prediction scenario has a lower probability value.

$$p = \begin{cases} p_{\mathrm{NP}}, & for\ \mathrm{NP} \\ p_{\mathrm{IP}}, & for\ \mathrm{IP} \\ p_{\mathrm{CP}}, & for\ \mathrm{CP} \end{cases} \tag{8}$$

where $p_{\mathrm{NP}} = p_{\mathrm{IP}} > p_{\mathrm{CP}}$.

By considering all four cost elements, the holistic cost function in (1) becomes:

$$C_{\mathrm{HO}} = \begin{cases} p_{\mathrm{NP}} C_{\mathrm{D}} + (1 - p_{\mathrm{NP}}) \times \\ \left[C_{\mathrm{L,NP}} + C_{\mathrm{OH,NP}} + C_{\mathrm{W,NP}}\right], & for\ \mathrm{NP} \\ p_{\mathrm{IP}} C_{\mathrm{D}} + (1 - p_{\mathrm{IP}}) \times \\ \left[C_{\mathrm{L,IP}} + C_{\mathrm{OH,IP}} + C_{\mathrm{W,IP}}\right], & for\ \mathrm{IP} \\ p_{\mathrm{CP}} C_{\mathrm{D}} + (1 - p_{\mathrm{CP}}) \times \\ \left[C_{\mathrm{L,CP}} + C_{\mathrm{OH,CP}} + C_{\mathrm{W,CP}}\right] & for\ \mathrm{CP}. \end{cases} \tag{9}$$

As all cost parameters have different scales, their effect on the holistic cost function may be different. This will result in the dominance of some elements because of their larger scale, while some are neglected due to their small scale. Thus, feature scaling (in the range of [0, 1]) is applied to the cost elements in order to keep their impact on the same scale. The feature scaling function is given by:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values in the set of x.

Hence, the following equations are obtained if (10) is applied to (4), (6), and (7), respectively. Note that feature scaling is not applied to (8), as it is already in the range [0, 1].

The HO latency is:

$$C'_{\mathrm{L}} = \begin{cases} \frac{C_{\mathrm{L,NP}} - C_{\mathrm{L,CP}}}{C_{\mathrm{L,IP}} - C_{\mathrm{L,CP}}}, & for\ \mathrm{NP}(\$C'_{\mathrm{L,NP}}\$) \\ 1, & for\ \mathrm{IP}(\$C'_{\mathrm{L,IP}}\$) \\ 0, & for\ \mathrm{CP}(\$C'_{\mathrm{L,CP}}\$). \end{cases} \tag{11}$$

Then, the signalling cost is:

$$C'_{\mathrm{OH}} = \begin{cases} \frac{C_{\mathrm{OH,NP}} - C_{\mathrm{OH,CP}}}{C_{\mathrm{OH,IP}} - C_{\mathrm{OH,CP}}}, & for\ \mathrm{NP}(\$C'_{\mathrm{OH,NP}}\$) \\ 1, & for\ \mathrm{IP}(\$C'_{\mathrm{OH,IP}}\$) \\ 0, & for\ \mathrm{CP}(\$C'_{\mathrm{OH,CP}}\$). \end{cases} \tag{12}$$

Finally, the resource wastage is given by

$$C'_{\mathrm{W}} = \begin{cases} 0, & for\ \mathrm{NP}(\$C'_{\mathrm{W,NP}}\$) \\ 1, & for\ \mathrm{IP}(\$C'_{\mathrm{W,IP}}\$) \\ \frac{C_{\mathrm{W,CP}} - C_{\mathrm{W,NP}}}{C_{\mathrm{W,IP}} - C_{\mathrm{W,NP}}}, & for\ \mathrm{CP}(\$C'_{\mathrm{W,CP}}\$). \end{cases} \tag{13}$$

Overall, the holistic cost function in (9) should be rearranged with feature scaled parameters:

$$C'_{\mathrm{HO}} = \begin{cases} p_{\mathrm{NP}} C_{\mathrm{D}} + \left(1 - p_{\mathrm{NP}}\right) \times \\ \left[\frac{C_{\mathrm{L,NP}} - C_{\mathrm{L,CP}}}{C_{\mathrm{L,IP}} - C_{\mathrm{L,CP}}} + \frac{C_{\mathrm{OH,NP}} - C_{\mathrm{OH,CP}}}{C_{\mathrm{OH,IP}} - C_{\mathrm{OH,CP}}}\right], & for\ \mathrm{NP} \\ p_{\mathrm{IP}} C_{\mathrm{D}} + \left(1 - p_{\mathrm{IP}}\right) \times 3, & for\ \mathrm{IP} \\ p_{\mathrm{CP}} C_{\mathrm{D}} + \left(1 - \frac{p_{\mathrm{CP}}}{p_{\mathrm{NP}} - p_{\mathrm{CP}}}\right) \times \\ \frac{C_{\mathrm{W,CP}} - C_{\mathrm{W,NP}}}{C_{\mathrm{W,IP}} - C_{\mathrm{W,NP}}}. & for\ \mathrm{CP} \end{cases} \tag{14}$$

Eq. (14) can be expressed as follows after plugging the corresponding cost values:

$$C'_{\mathrm{HO}} = \begin{cases} p_{\mathrm{NP}} C_{\mathrm{D}} + 1.2\left(1 - p_{\mathrm{NP}}\right), & for\ \mathrm{NP} \\ p_{\mathrm{IP}} C_{\mathrm{D}} + 3\left(1 - p_{\mathrm{IP}}\right), & for\ \mathrm{IP} \\ p_{\mathrm{CP}} C_{\mathrm{D}} + \frac{t_{\mathrm{a}}}{t_{\mathrm{m}}}\left(1 - p_{\mathrm{CP}}\right), & for\ \mathrm{CP}. \end{cases} \tag{15}$$

It can be seen from (15) that $C'_{\mathrm{HO,CP}} < C'_{\mathrm{HO,NP}} < C'_{\mathrm{HO,IP}}$ where $C'_{\mathrm{HO,NP}}$, $C'_{\mathrm{HO,CP}}$, and $C'_{\mathrm{HO,IP}}$, are feature scaled versions of the holistic costs for no-prediction, correct and incorrect predictions, respectively.

Since the predictive process includes both the correct and incorrect predictions, the total cost of a predictive process is required to include both; i.e.

$$C_{\mathrm{HO,P}} = A C_{\mathrm{HO,CP}} + (1 - A) C_{\mathrm{HO,IP}} \tag{16}$$

where A is the prediction accuracy that can be defined as follows:

$$A = \frac{N_{\mathrm{p}}^{\mathrm{c}}}{N_{\mathrm{p}}^{\mathrm{c}} + N_{\mathrm{p}}^{\mathrm{i}}} \tag{17}$$

where $N_{\mathrm{p}}^{\mathrm{c}}$ and $N_{\mathrm{p}}^{\mathrm{i}}$ are the number of correct and incorrect predictions, respectively.

Alternatively, the feature scaled version of (16) is:

$$C'_{\mathrm{HO,P}} = A C'_{\mathrm{HO,CP}} + (1 - A) C'_{\mathrm{HO,IP}} \tag{18}$$

The following expression can be derived if (15), (17), and (18) are combined:

$$C'_{\mathrm{HO,P}} = \frac{N_{\mathrm{p}}^{\mathrm{c}}}{N_{\mathrm{p}}^{\mathrm{c}} + N_{\mathrm{p}}^{\mathrm{i}}} \times \left[p_{\mathrm{CP}} C_{\mathrm{D}} + \left(1 - p_{\mathrm{CP}}\right) \times \frac{t_{\mathrm{a}}}{t_{\mathrm{m}}}\right] + \left(1 - \frac{N_{\mathrm{p}}^{\mathrm{c}}}{N_{\mathrm{p}}^{\mathrm{c}} + N_{\mathrm{p}}^{\mathrm{i}}}\right) \times \left[p_{\mathrm{IP}} C_{\mathrm{D}} + \left(1 - p_{\mathrm{IP}}\right) \times 3\right]. \tag{19}$$

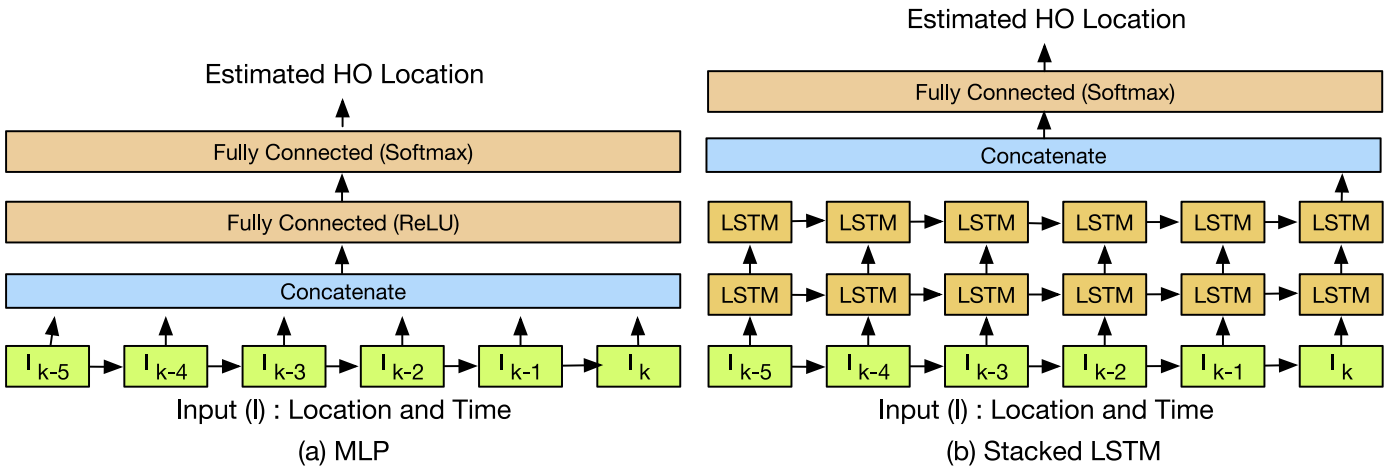

Fig. 3. (a) Multi-layered perceptron and (b) Stacked long-short-term memory based HO prediction model: Example of 5 prior input location and time (taking into account the current location I_k as well as temporal information of 5 previous inputs I_{k-1}, I_{k-2}, I_{k-3}, I_{k-4}, I_{k-5}).

As the predictive process is employed to perform better than the conventional non-predictive process, the overall cost in (19) is supposed to be lower than the overall cost of the non-predictive process. Accordingly, this constraint can be modelled as follows:

$$C'_{\mathrm{HO,P}} \leq C'_{\mathrm{HO,NP}}. \tag{20}$$

However, this argument would change according to network conditions. For resource-intolerant networks, for example, the resource wastage is priority and cannot be sacrificed. We will revert to this issue later in this section.

In order to find the minimum accuracy ($A_{\min}$) satisfying the constraint (20), the equality of $C'_{\mathrm{HO,P}}$ and $C'_{\mathrm{HO,NP}}$ should be considered; such that:

$$\begin{aligned} &C'_{\mathrm{HO,P}} = C'_{\mathrm{HO,NP}}, \\ &A_{\min} C'_{\mathrm{HO,CP}} + (1 - A_{\min}) C'_{\mathrm{HO,IP}} = C'_{\mathrm{HO,NP}}, \\ &A_{\min} = \frac{C'_{\mathrm{HO,NP}} - C'_{\mathrm{HO,IP}}}{C'_{\mathrm{HO,CP}} - C'_{\mathrm{HO,IP}}}. \end{aligned} \tag{21}$$

In other words, any predictor, which is to be employed in a predictive HO process, should satisfy at least the accuracy level in (21) to meet the constraint in (20). Moreover, it is obvious from (16) and (18) that the only way to decrease $C'_{\mathrm{HO,P}}$ is to increase the accuracy; i.e. the more the accuracy, the less the HO cost $C'_{\mathrm{HO,P}}$. Therefore, since the accuracy level plays a vital role, the focus of the predictive HO scheme should be on increasing the accuracy level as much as possible. To this end, it is worth investigating appropriate methods to boost the accuracy of predictors.

In addition, prioritization can also be taken into account by appending corresponding parameters to the derived equations. These parameters are responsible for prioritizing the elements in (1); such that

$$\ddot{C}_{\mathrm{HO}} = p\alpha C_{\mathrm{D}} + (1-p)[\beta C_{\mathrm{L}} + \gamma C_{\mathrm{OH}} + \zeta C_{\mathrm{W}}], \tag{22}$$

where $\ddot{C}_{\mathrm{HO}}$ is an element-prioritized version of C_{HO}; $\{\alpha, \beta, \gamma, \zeta\} \in \mathbb{R}$ are the prioritization parameters for user dissatisfaction, latency, signalling overhead, and resource wastage, respectively. For ultra-dense networks, for example, the resource wastage can be the most important element, as the network can sacrifice user satisfaction in order to accommodate more users. On the contrary, it can be the least crucial element in rural areas, as there is more likely to be an abundance of resources due to the limited number of users. In other words, these prioritization parameters enable networks to prioritize any element(s) based on the circumstances they experience. Therefore, the minimum accuracy requirement in (21) is subject to an alteration based on the prioritization parameters introduced in (22).

5. Data-driven HO prediction

This section describes the proposed machine learning-driven HO prediction algorithms, to further minimize the cost function and improve mobility management. Specifically, shallow MLP and deep stacked LSTM learning based HO prediction approaches are described. Note that different input combinations, hidden neurons, and LSTM cells have been explored to achieve the best possible HO prediction.

5.1. MLP based HO prediction

This section describes the three-layered feed-forward MLP network architecture summarized in Fig. 3(a). The network consists of an input layer, a hidden layer with rectified linear activation functions, and a softmax output layer. The network was trained using the stochastic gradient-decent and Adam optimizer to minimize the categorical cross-entropy. For training, different number of hidden neurons ranging from 16 to 128 were used. Comparative simulation results showed that, when the number of neurons in the hidden layer increase, the validation accuracy increases. However, further increase in the number of hidden neurons led to network overfitting on the training data, thus leading to poor generalization on the validation set. For the proposed HO prediction model, 64 hidden neurons provided the most optimal results. More details are presented in subsequent sections.

5.2. Stacked LSTM based HO prediction

This section describes the LSTM network architecture summarised in Fig. 3(b). The LSTM network consists of an input layer, two stacked LSTM layers followed by a softmax layer. Inputs (Time and Locations) of time instance I_k, I_{k-1},..., I_{k-5} (where k is the current time and 5 is the number of prior inputs) are fed into stacked LSTM layers. In the network architecture, the input at layer k is the value of hidden state h_t computed by layer $k-1$. The lower LSTM has 32 cells, which encode the input and passes on hidden states to the second LSTM layer with 32 cells. The output of the second LSTM is fed into a fully connected layer which has 20 neurons, with softmax activation. The architecture was trained to minimize the categorical cross-entropy using stochastic gradient-decent algorithm and Adam optimizer. Note that previous works

Table 1
Structure of the data.

x	y	date and time (dd/mm/yyyy hh.mm)	ID

Table 2
Structure of the data after pre-processing.

date and time (dd/mm/yyyy)	cell	ID

[27,28] have shown that the stacked LSTM, due to its inherent recurrent nature, can better model long-term dependencies and exploit temporal correlation in inputs, as compared to the MLP.

6. Performance evaluation

HO prediction is treated as a classification problem in which each DBS constitutes one class. Hence, the task is to predict the next location of the user given the current time and previous locations visited. The MLP and stacked LSTM based data-driven approaches for HO prediction are trained, validated and compared using the PyTorch library. The modeled MLP/LSTM engines learn patterns belonging to the input layer and classify them according to their serving DBSs. The proposed network architectures and employed parameters were presented in Section 5. Next, the benchmark dataset used in this study is described.

6.1. Data set

The data set was developed by MIT Human Dynamics Lab [29], based on social studies conducted on human behaviours and interactions. The data belongs to a data server configuration company at Chicago, and was collected from employees of the company [30]. Each participating employee was given a badge with a unique ID number, and parameters such as location, audio intensity, etc., were collected through the distributed badges. Locations of users were estimated via Zigbee RSSI (Received Signal Strength Indicator) readings using an algorithm detailed in [30], where the locations of employees were mapped to 502 evenly distributed grids with specific x and y coordinates. Furthermore, the data consists of x and y coordinates of grids, where the employees are located, along with corresponding time and dates, as presented in Table 1. In total, 1,006,190 sequences belonging to 39 IDs are available. Sampling frequency is provided in data description on the dataset website [31], specifically: "10 locations with longest stays *(x, y)* per employee (identified by the id of the badge assigned to the employee) per minute (time), estimated from Zigbee RSSI to anchor nodes." Note that the amount of samples is not the same for each ID, and there are also missing entries throughout the data set.

6.1.1. Data preprocessing

We first selected 11 IDs, whose number of sequences are provided in Table 3. The pre-processed data was split into Train (70%), Validation (10%) and Test (20%) sets to facilitate this study. Specifically, we down-sampled the data by creating 20 square-shaped cells based on x and y coordinates, and labelled them from 1 to 20 sequentially. Next, each user was associated with one grid that reflected the corresponding cell. In other words, locations of users with x and y coordinates are mapped to their respective cells. In addition, the locations were encoded into a 20-D vector. Both LSTM and MLP employ inputs (time and locations) comprising time instances I_k, I_{k-1},..., I_{k-n} as shown in Fig. 3, where n is the number of prior inputs and k is the current time step. Table 2 shows the structure of data after pre-processing.

Table 3
Summary of train, validation, and test samples for each user.

User ID	Train	Validation	Test
56	10080	1440	2880
82	14979	2140	4280
99	18774	2682	5364
101	12264	1752	3504
106	30428	4347	8694
109	22232	3176	6352
266	17010	2430	4860
268	21987	3141	6282
280	30358	4337	8674
291	7741	1106	2212
297	12467	1781	3562

Table 4
Parameters for numerical simulations.

Parameter	Value	Description
p_{NP}	0.6	Probability of no enough resource (NP)
p_{IP}	0.6	Probability of no enough resource (IP)
p_{CP}	0.3	Probability of no enough resource (CP)
C_D	1	Dissatisfaction cost
t_a/t_m	1/5	Ratio of advance reservation and maximum waiting times
β, γ, ζ	1	Prioritization parameters

The motivation for this pre-processing is to exploit the temporal correlation between past locations and to make the data suitable for the classification problem, such that each cell constitutes a class. Moreover, the computational cost is also reduced as the number of output classes is reduced from 502 to 20.

6.2. Numerical results

Numerical results related to prediction accuracy and the holistic cost are first presented in order to establish the importance of the accuracy level for predictive HO management. The parameters used for numerical simulation are depicted in Table 4.

The numerical results for various α are shown in Fig. 4. The intersection points of predictive and non-predictive processes reflect the minimum required accuracy, since the holistic cost exhibited by the predictive and the non-predictive approach is equal. As shown in Fig. 4, lower values of α decrease the minimum required accuracy (around 35% for $\alpha = 0.1$) and also the holistic cost. On the other hand, higher values of α increase the required minimum accuracy level (more than 80% for $\alpha = 10$). The aforementioned results imply that the accuracy level plays a vital role in predictive HO management, and prioritization parameters are key to determine the minimum accuracy level.

Consequently, the holistic cost and minimum accuracy requirement with prioritization parameters can benefit the designing of a predictive HO management system, as summarized below:

- The holistic cost and minimum accuracy requirement can help jointly determine the applicability of the designed predictor.
- The prioritization parameters contribute to organizing the holistic HO cost based on conditions experienced in the network.
- The minimum accuracy requirement information renders the design process more formal and effective, as predictors can be chosen according to the minimum accuracy requirement and computational cost.

6.3. Experimental results

For HO prediction, both MLP and LSTM based data-driven approaches are used. First, a MLP based learning model is trained to

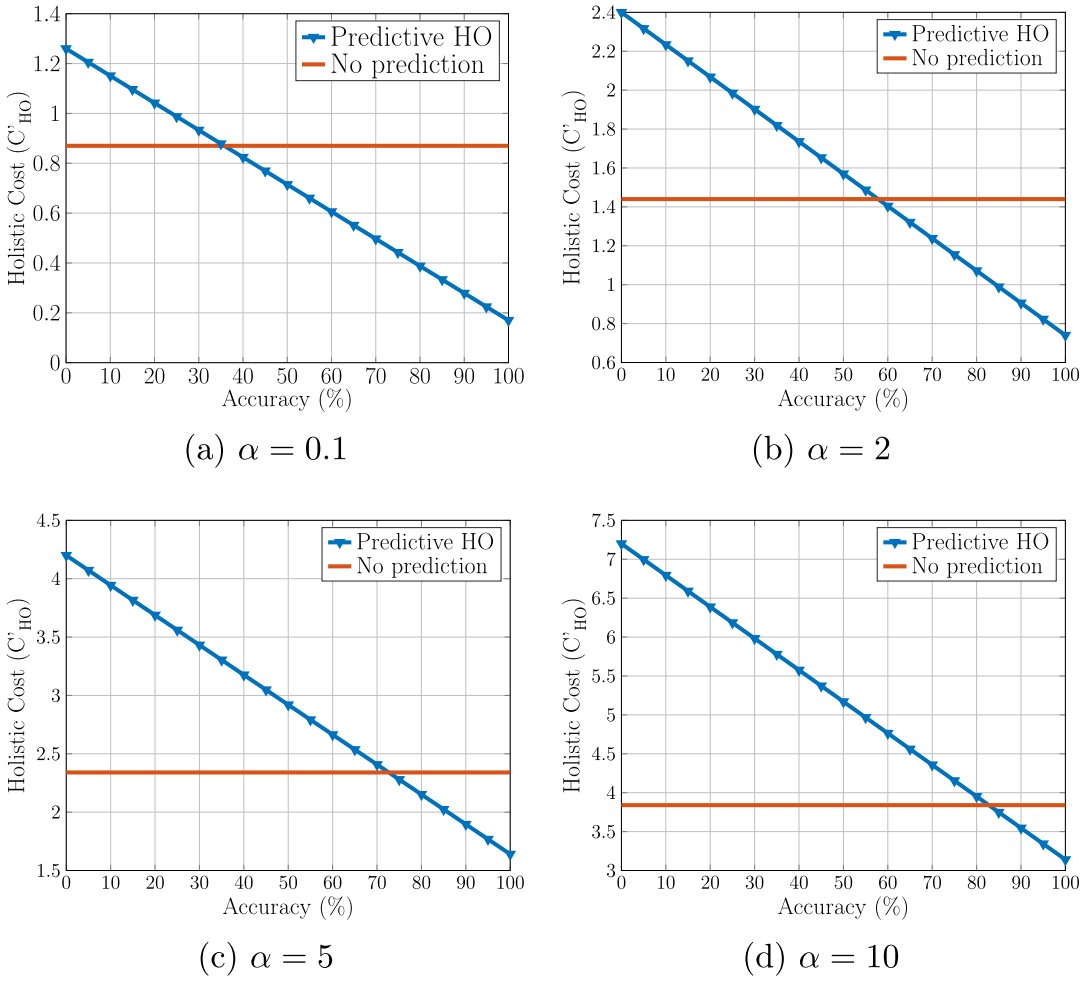

Fig. 4. Numerical results for various α values. Holistic cost is calculated through (15), which includes user dissatisfaction; HO latency; overhead; and resource wastage, and using values in Table 4.

identify the optimal number of hidden neurons, ranging from 8 to 64. In addition, a LSTM based HO model is trained to identify the optimal number of cells in the LSTM layer, ranging from 16 to 128. Fig. 5a demonstrates the prediction performance of MLP for four different number of hidden layer units (ζ), while Fig. 5b shows the prediction performance of LSTM for three different number of LSTM cells (η).

Before analyzing the obtained results, we evaluate the minimum accuracy requirement. Using (21) with the values given in Table 4, and setting α to 1, the required minimum prediction accuracy, $A_{\min}$, becomes 52.9%.

Therefore, we conclude that any predictive system must achieve an accuracy greater than 52.9% in order to outperform the non-predictive system.

The results in Fig. 5 reveal that although ζ and η seem to have little impact on the prediction performances of MLP and LSTM, the validation accuracy increases as the number of hidden neurons and LSTM size is increased. However, further increase in these parameters leads to overfitting, resulting in poor validation accuracy. In addition, it is worth noting that the number of hidden neurons and LSTM size play a significant role in the deployment of these algorithms at large scale. We found $\zeta = 64$ and $\eta = 32$ are optimal for the considered dataset. Moreover, for ID 291, the MLP shows a 7% validation accuracy improvement when ζ is changed form 16 to 64, and the LSTM shows a 6% validation accuracy improvement when η is changed from 8 to 32. The complexity of a neural network can be assessed by analyzing the number of parameters in the network. The total number of parameters in MLP and LSTM networks were found to equal 16276 and 16424, respectively. Therefore, it can be concluded that both MLP and LSTM have similar computational complexity.

Table 5
MLP and LSTM validation accuracy (%) for various input time-instances (I_k, I_{k-1},..., I_{k-n}); e.g., 4 corresponds to input comprising the current time instance and 3 prior time instances.

User ID	MLP				LSTM			
	1	4	8	12	1	4	8	12
297	14.39	30.52	31.84	30.86	17.95	29.63	28.93	27.97
101	25.08	27.72	31.07	31.79	30.08	30.83	31.34	30.75
268	24.51	32.92	30.76	31.46	32.63	33.01	31.02	30.07
99	38.81	41.69	44.01	44.44	40.47	41.93	45.21	36.31
266	38.93	44.18	50.75	49.05	38.59	49.97	48.95	48.79
56	43.94	46.13	46.55	46.25	44.63	43.99	49.58	47.95
280	43.90	42.17	47.84	47.66	47.96	47.53	52.48	52.37
291	53.24	56.06	60.89	59.59	55.95	57.61	61.58	60.61
82	53.15	61.87	65.21	64.79	64.02	66.96	66.84	67.80
109	65.68	71.18	73.47	72.63	73.49	75.59	77.95	76.63
106	82.69	82.08	83.98	83.78	84.08	83.67	85.34	85.76

In multiple, prior-location to next-location mapping, various input time instances (I_k, I_{k-1},..., I_{k-n}) ranging from 1 to 12 are used (e.g., 4 corresponds to input comprising the current time instance and 3 prior time instances). The results for different number of prior inputs are presented in Table 5. It can be seen that the

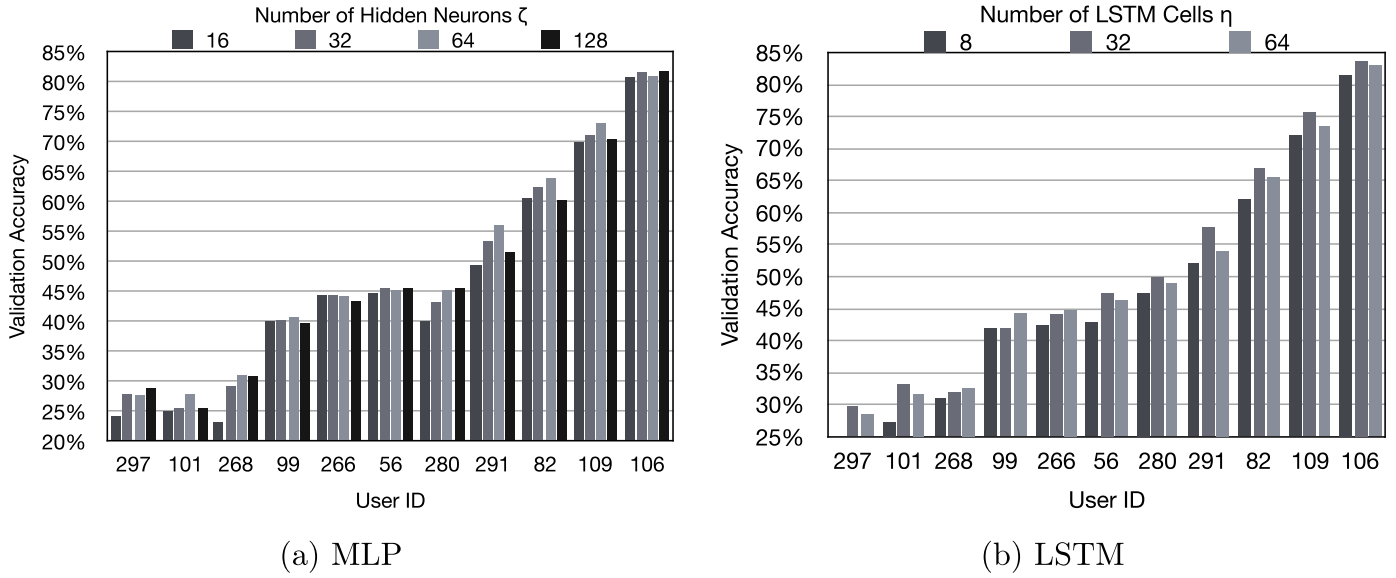

Fig. 5. Validation accuracy for both MLP (5a) and LSTM (5b). Users IDs (56, 82, 99, 101, 106, 109, 266, 268, 280, 291, 297) are given by the data set provider. ζ and η are the number of hidden layer units and LSTM size, respectively.

Table 6
Test accuracy (%) results of MLP and LSTM for (a) Time only input and (b) Time and Location input.

User ID	Time only		Time & Location	
	MLP	LSTM	MLP	LSTM
297	17.84	18.16	27.23	31.66
101	23.74	22.03	30.59	34.45
268	19.15	19.87	30.99	36.03
99	29.98	29.98	40.87	43.91
56	15.63	16.77	46.11	47.23
280	37.79	40.55	47.78	49.30
266	34.99	34.99	45.92	53.88
291	57.58	56.58	55.63	59.58
82	18.05	22.13	63.82	65.32
109	64.75	64.75	70.44	73.96
106	78.28	78.30	82.45	84.56

addition of prior temporal information improves the overall network accuracy for both the LSTM and MLP models. However, further addition of prior temporal information leads to inclusion of unnecessary information in the input, thus resulting in poor validation accuracy. We found 8 input time instances (i.e. current location and 7 previous locations) to be optimal for the given dataset.

Optimal tuning parameters for the validation set were used to train the MLP and LSTM for two cases: (a) *Time only* input and (b) *Time and Location* input. Table 6 depicts test accuracy results for both cases. *Time & Location* significantly outperforms *Time only* for all users. Even though the LSTM and MLP have similar computational complexity, the LSTM outperforms MLP for all users, achieving up to 8% performance improvement for User 266.

7. Conclusion

In this work, we first developed a novel analytical model to simultaneously meet the holistic cost and minimum prediction accuracy requirements. The holistic cost model integrates signaling overhead, latency, call dropping, and radio resource wastage. Further, novel data-driven deep learning approaches were employed to complement the holistic model and further reduce the holistic cost. The evaluation of the holistic model at different prioritization parameters (user dissatisfaction, latency, signalling overhead, and resource wastage) demonstrated the impact of prediction accuracy as part of the overall cost. We conclude that our developed model coupled with a data-driven, deep learning approach can lead to reduced user dissatisfaction, latency, signalling overhead, and resource wastage. The innovative model is thus posited as benchmark resource for the mobile networking and machine learning communities.

Future work includes the development of context-aware switching between non-predictive and predictive models and optimal tuning of prioritization parameters (α, β, γ, and ζ), according to network conditions and requirements, using state-of-the-art optimization algorithms, such as reinforcement learning [32,33] and nature-inspired approaches [34,35]. This would enable the network to perform dynamic parameter tuning, and further enhance both the network performance and user satisfaction in dynamic scenarios. In addition, ongoing work is aimed at carrying out further extensive evaluation with larger and more complex real-world data sets, considering required trade-offs between computational cost and real-time decision-making.

Acknowledgements

A. Hussain and A. Ahsan were supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant no. EP/M026981/1. M. A. Imran and O. Onireti were supported by EPSRC Global Challenges Research Fund (GCRF) DARE project EP/P028764/1. M. Ozturk is supported by the Republic of Turkey Ministry of National Education (MoNE-1416/YLSY). CBS, DBS, and UE icons in Figs. 1 and 2 were made by Freepik, Kiranshastry, and Smashicons from www.flaticon.com, respectively. A. Hussain would like to acknowledge helpful discussions with A. Hawalah from Taibah Valley, at Taibah University (Madinah, Saudi Arabia).

References

[1] O.G. Aliu, A. Imran, M.A. Imran, B. Evans, A survey of self organisation in future cellular networks, IEEE Commun. Surv. Tutor. 15 (1) (2013) 336–361.
[2] A. Mohamed, O. Onireti, M.A. Imran, A. Imran, R. Tafazolli, Control-data separation architecture for cellular radio access networks: a survey and outlook, IEEE Commun. Surv. Tutor. 18 (1) (2016) 446–465.
[3] A. Nadembega, A. Hafid, T. Taleb, A destination and mobility path prediction scheme for mobile networks, IEEE Trans. Veh. Technol. 64 (6) (2015) 2577–2590.

[4] A. Mohamed, O. Onireti, M.A. Imran, A. Imran, R. Tafazolli, Predictive and core-network efficient RRC signalling for active state handover in RANs with control/data separation, IEEE Trans. Wirel. Commun. 16 (3) (2017) 1423–1436.
[5] A. Mohamed, O. Onireti, S.A. Hoseinitabatabaei, M. Imran, A. Imran, R. Tafazolli, Mobility prediction for handover management in cellular networks with control/data separation, in: Proceedings of the IEEE International Conference on Communications (ICC), IEEE, 2015, pp. 3939–3944.
[6] M. Ozturk, P.V. Klaine, M.A. Imran, 3D transition matrix solution for a path dependency problem of Markov Chains-based prediction in cellular networks, in: Proceedings of the IEEE Eighty-Sixth Vehicular Technology Conference (VTC–Fall), IEEE, 2017, pp. 1–5.
[7] H. Abu-Ghazaleh, A.S. Alfa, Application of mobility prediction in wireless networks using Markov renewal theory, IEEE Trans. Veh. Technol. 59 (2) (2010) 788–802.
[8] M. Ozturk, P.V. Klaine, M.A. Imran, Improvement on the performance of predictive handover management by setting a threshold, in: Proceedings of the IEEE Eighty-Sixth Vehicular Technology Conference (VTC-Fall), IEEE, 2017, pp. 1–5.
[9] A. Quintero, A user pattern learning strategy for managing users' mobility in UMTS networks, IEEE Trans. Mob. Comput. 4 (6) (2005) 552–566.
[10] N. Sinclair, D. Harle, I.A. Glover, J. Irvine, R.C. Atkinson, An advanced SOM algorithm applied to handover management within LTE, IEEE Trans. Veh. Technol. 62 (5) (2013) 1883–1894.
[11] A. Quintero, O. Garcia, A profile-based strategy for managing user mobility in third-generation mobile systems, IEEE Commun. Mag. 42 (9) (2004) 134–139.
[12] S. Goudarzi, W.H. Hassan, M.H. Anisi, A. Soleymani, M. Sookhak, M.K. Khan, A.-H.A. Hashim, M. Zareei, Abc-pso for vertical handover in heterogeneous wireless networks, Neurocomputing 256 (2017) 63–81.
[13] Y. Qiao, Z. Si, Y. Zhang, F.B. Abdesslem, X. Zhang, J. Yang, A hybrid Markov-based model for human mobility prediction, Neurocomputing 278 (2018) 99–109.
[14] X. Yang, K. Huang, R. Zhang, A. Hussain, Learning latent features with infinite nonnegative binary matrix trifactorization, IEEE Trans. Emerg. Top. Comput. Intell. 2 (5) (2018) 450–463.
[15] F. Xiong, B. Sun, X. Yang, H. Qiao, K. Huang, A. Hussain, Z. Liu, Guided policy search for sequential multitask learning, IEEE Trans. Syst. Man Cybern. Syst. 49 (1) (2018) 216–226.
[16] A. Adeel, H. Larijani, A. Ahmadinia, Random neural network based novel decision making framework for optimized and autonomous power control in LTE uplink system, Phys. Commun. 19 (2016) 106–117.
[17] G. Aceto, D. Ciuonzo, A. Montieri, A. Pescapé, Mobile encrypted traffic classification using deep learning, in: Proceedings of the IEEEACM Network Traffic Measurement and Analysis Conference (TMA 2018), 2018.
[18] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, A.M. Dobaie, Facial expression recognition via learning deep sparse autoencoders, Neurocomputing 273 (2018) 643–649.
[19] N. Zeng, Z. Wang, H. Zhang, W. Liu, F.E. Alsaadi, Deep belief networks for quantitative analysis of a gold immunochromatographic strip, Cogn. Comput. 8 (4) (2016) 684–692.
[20] T. Zhang, E. van den Berg, J. Chennikara, P. Agrawal, J.-C. Chen, T. Kodama, Local predictive resource reservation for handoff in multimedia wireless ip networks, IEEE J. Sel. Areas Commun. 19 (10) (2001) 1931–1941.
[21] X. Luo, I. Thng, W. Zhuang, A dynamic channel pre-reservation scheme for handoffs with GoS guarantee in mobile networks, in: Proceedings of the IEEE International Symposium on Computers and Communications, IEEE, 1999, pp. 404–408.
[22] B.M. Epstein, M. Schwartz, Predictive Qos-based admission control for multiclass traffic in cellular wireless networks, IEEE J. Sel. Areas Commun. 18 (3) (2000) 523–534.
[23] K. Alexandris, N. Nikaein, R. Knopp, C. Bonnet, Analyzing x2 handover in lte/lte-a, in: Proceedings of the Fourteenth International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), IEEE, 2016, pp. 1–7.
[24] 3GPP, Evolved universal terrestrial radio access (E -UTRA); Requirements for support of radio resource management, TS 36.133, 3rd Generation Partnership Project (3GPP), 2009. URL https://www.arib.or.jp/english/html/overview/doc/STD-T104v1_10/5_Appendix/Rel10/36/36133-a40.pdf.
[25] J.S. Ho, I.F. Akyildiz, Local anchor scheme for reducing signaling costs in personal communications networks, IEEE/ACM Trans. Netw. (TON) 4 (5) (1996) 709–725.
[26] J. Martinez-Bauset, J.M. Gimenez-Guzman, V. Pla, Optimal admission control in multimedia mobile networks with handover prediction, IEEE Wirel. Commun. 15 (5) (2008) 38–44.
[27] A. Adeel, J. Ahmad, A. Hussain, Real-time lightweight chaotic encryption for 5G-IoT enabled lip-reading driven secure hearing-aid, arXiv:1809.04966, (2018).
[28] M. Gogate, A. Adeel, R. Marxer, J. Barker, A. Hussain, DNN-driven speaker independent audio-visual mask estimation for speech separation, in: Proceedings of Interspeech 2018, 2018, pp. 2723–2727.
[29] D.O. Olguín, B.N. Waber, T. Kim, A. Mohan, K. Ara, A. Pentland, Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior, IEEE Trans. Syst. Man Cybernet. Part B (Cybernetics) 39 (1) (2009) 43–55.
[30] W. Dong, D. Olguin-Olguin, B. Waber, T. Kim, et al., Mapping organizational dynamics with body sensor networks, in: Proceedings of the Ninth International Conference on Wearable and Implantable Body Sensor Networks (BSN), IEEE, 2012, pp. 130–135.
[31] MIT human dynamics lab - reality commons, (http://realitycommons.media.mit.edu/badgedataset1.html). Accessed: 2018-19-10.
[32] M. Mahmud, M.S. Kaiser, A. Hussain, S. Vassanelli, Applications of deep learning and reinforcement learning to biological data, IEEE Trans. Neural Networks Learn. Syst. 29 (6) (2018) 2063–2079.
[33] P.V. Klaine, J.P.B. Nadas, R.D. Souza, M.A. Imran, Distributed drone base station positioning for emergency cellular networks using reinforcement learning, Cogn. Comput. 10 (5) (2018) 790–804.
[34] X.S. Yang, S. Deb, S.K. Mishra, Multi-species cuckoo search algorithm for global optimization, Cogn. Comput. 10 (6) (2018) 1085–1095.
[35] W.A.H.M. Ghanem, A. Jantan, A cognitively inspired hybridization of artificial bee colony and dragonfly algorithms for training multi-layer perceptrons, Cogn. Comput. 10 (6) (2018) 1096–1134.

Metin Ozturk received the B.Sc. degree in electrical and electronics engineering from Eskisehir Osmangazi University, Turkey, in 2013, and the M.Sc. degree in electronics and communication engineering from Ankara Yildirim Beyazit University, Turkey, in 2016. He also served as a research assistant at the latter university from 2013 to 2016. He is currently pursuing his Ph.D. degree with the School of Engineering, University of Glasgow. His research interests include Self-Organizing Networks (SON), mobile communications, and machine learning algorithms.

Mandar Gogate obtained his B.Eng. in Electronics (with the highest 1st Class Honours with distinction) from BITS Pilani, India, in 2016. During 2015-16, he worked as a Research assistant at ENSTA ParisTech - École Nationale SupĘrieure de Techniques AvancĘes, Paris, France where he researched deep learning models for Multimodal Robotic sensor fusion and Incremental learning. He is currently a full-time Ph.D. researcher in the Cognitive Big Data and Cybersecurity (CogBiD) Research Lab at Edinburgh Napier University, in Scotland, UK. He is working on multimodal big data fusion using deep and incremental learning, for solving a number of challenging real-world problems, including cybersecurity, speech separation, sentiment and opinion mining and 5G-IoT applications.

Oluwakayode Onireti received the B.Eng. degree in electrical engineering from the University of Ilorin, Ilorin, Nigeria, in 2005, and the M.Sc. degree in mobile and satellite communications, and the Ph.D. degree in electronics engineering from the University of Surrey, Guildford, U.K., in 2009 and 2012, respectively. He secured a first class grade in his B.Eng. and a distinction in his M.Sc. degree. From 2013 to 2016, he was a Research Fellow with the Institute for Communication Systems, University of Surrey, Guildford, U.K. He is currently a Research Associate with the School of Engineering, University of Glasgow, Glasgow, U.K. He has been actively involved in projects, such as ROCKET, EARTH, Greencom, QSON, and Energy proportional EnodeB for LTE-Advanced and Beyond. He is currently involved in the DARE project, a ESPRC funded project on distributed autonomous and resilient emergency management systems. His main research interests include self-organizing cellular networks, energy efficiency, multiple-input multiple-output, and cooperative communications.

Ahsan Adeel holds B. Eng. (Hons), MSc (EEE), and PhD (Cognitive Computing) degrees. Following an EPSRC/MRC prestigious fellowship at the University of Stirling (2016-18), he is currently a Lecturer (Assistant Professor) in Computing Science at the University of Wolverhampton, UK, where he is leading the Conscious Multisensory Integration (CMI) Lab. He is a Visiting Fellow at MIT Synthetic Intelligence Lab and Computational Neuroscience Lab (University of Oxford). His ongoing multidisciplinary research aims to explore and exploit the power of advanced AI to design unorthodox brain-inspired cognitive computing architectures by integrating suitable deep machine learning, reasoning, and optimization algorithms. His focused approaches include biophysical and hardware-efficient neural models, explainable artificial intelligence, optimized resource management, multimodal fusion, context-aware decision-making, low power 5G IoT devices, and neuromorphic chips.

Amir Hussain obtained his B.Eng. (with the highest 1st Class Honors) and Ph.D. (in novel neural network architectures and algorithms) from the University of Strathclyde in Glasgow, Scotland, UK, in 1992 and 1997 respectively. Following postdocoral and academic positions at the University of West of Scotland (1996-98), University of Dundee (1998-2000), and University of Stirling (2000-2018) respectively, he joined Edinburgh Napier University, in Scotland, UK, in 2018, as Professor of Computing Science, and founding Director of the Cognitive Big Data and Cybersecurity (CogBiD) Research Laboratory. His research interests are cross-disciplinary and industry focussed, and include secure and context-aware 5G-IoT driven AI, and multi-modal cognitive and sentic computing techniques and applications. He has published more than 400 papers, including over a dozen books and around 150 journal papers. He has led major national, European and international projects and supervised more than 30 PhD students, He is founding Editor-in-Chief of two leading journals: Cognitive Computation (Springer Nature), and BMC Big Data Analytics (BioMed Central); and Chief-Editor of the Springer Book Series on Socio-Affective Computing, and Cognitive Computation Trends. He has been appointed invited Associate Editor of several prestigeous journals, including the IEEE Transactions on Neural Networks and Learning Systems, the IEEE Transactions on Emerging Topics in Computational Intelligence, and (Elsevier) Information Fusion. He is Vice-Chair of the Emergent Technologies Technical Committee of the IEEE Computational Intelligence Society (CIS), and Chapter Chair of the IEEE UK and RI Industry Applications Society.

Muhammad Ali Imran Fellow IET, Senior Member IEEE, Senior Fellow HEA is a Professor of Wireless Communication Systems with research interests in self organised networks, wireless networked control systems and the wireless sensor systems. He heads the Communications, Sensing and Imaging (CSI) research group at the University of Glasgow. He is an Affiliate Professor at the University of Oklahoma, USA and a visiting Professor at 5G Innovation Centre, University of Surrey, UK. He has over 18 years of combined academic and industry experience with several leading roles in multi-million pound funded projects. He has been awarded 15 patents; has authored/coauthored over 400 journal and conference publications; was editor of 2 books and author of more than 15 book chapters; has successfully supervised over 40 postgraduate students at Doctoral level. He has been a consultant to international projects and local companies in the area of self-organised networks.