Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 155 (2019) 121–128

Procedia Computer Science

www.elsevier.com/locate/procedia

The 16th International Conference on Mobile Systems and Pervasive Computing (MobiSPC)
August 19-21, 2019, Halifax, Canada

Efficient Mobile Clouds: Forecasting the Future Connectivity of Mobile and IoT Devices to Save Energy and Bandwidth

Gabriel Orsini*, Wolf Posdorfer, Winfried Lamersdorf

[a] *Distributed Systems Group, Department of Computer Science, University of Hamburg, Germany*

Abstract

Use cases in the Internet of Things (IoT) and in mobile clouds often require the interaction of one or more mobile devices with their infrastructure to provide users with services. Ideally, this interaction is based on a reliable connection between the communicating devices, which is often not the case. Since most use cases do not adequately address this issue, service quality is often compromised.

Aimed to address this issue, this paper proposes a novel approach to forecast the connectivity and bandwidth of mobile devices by applying machine learning to the context data recorded by the various sensors of the mobile device. This concept, designed as a microservice, has been implemented in the mobile middleware *CloudAware*, a system software infrastructure for mobile cloud computing that integrates easily with mobile operating systems, such as Android.

We evaluate our approach with real sensor data and show how to enable mobile devices in the IoT to make assumptions about their future connectivity, allowing for intelligent and distributed decision making on the mobile edge of the network.

Keywords: Mobile Clouds; Internet of Things; Context Awareness; Context Forecast;

1. Introduction

Mobile devices like smartphones, wearables and sensor nodes have become more powerful every year. Nevertheless, they often rely on the resource augmentation through centralized resources, enabling a multitude of cloud-augmented mobile applications. Examples are location-based advertising [9], realtime sensor networks, the *Nvidia Shield* video-gaming console [13], which computes parts of the gameplay on remote resources, or the voice recognition assistant *Siri* [3]. Common to these use cases is the fact, that they rely on a preferably fast and stable connection to centralized or edge clouds [1]. Even with upcoming 5G networks it is believed that the obstacle of the intermittent connectivity of mobile devices will persist [17].

Knowledge about the future connectivity and bandwidth of mobile devices allows to design their applications accordingly and allow an improved usability and user experience, for example by deciding whether to prefetch data or postpone the synchronization with cloud services.

* Corresponding author. Tel.: +49-40-42883-2140 ; fax: +49-40-42883-2328.
E-mail address: [orsini,posdorfer,lamersd]@informatik.uni-hamburg.de

10.1016/j.procs.2019.08.020

Moreover, information about the current and future bandwidth can be used to decide when to activate the wireless network interfaces, e.g. GSM or WiFi, of the mobile device. Hereby, the interfaces will only be activated in situations where a high bandwidth can be achieved, improving the ratio between the required energy and the transferred data and thus saving energy, often the most limited resource on mobile devices.

Currently, many of the proposed solutions in the domain of connectivity forecasts are limited to specific use cases and only allow short-term forecasts. Filling this gap, this paper, which is based on our previous works in the domain of context forecasts [14, 15, 16], aims to provide a concept for the connectivity forecast on a multitude of different mobile devices, allowing them to reason about their future bandwidth. This way, a more efficient interaction between mobile devices as well as between cloud and edge cloud resources becomes possible and can provide a higher user experience while at the same time saving energy. Hence, the contributions in this paper can be summarized as follows:

- The concept of a modular connectivity service that is able to run on a multitude of devices. Firstly through the dynamic adaptation of the computational requirements and secondly through the adaptation to the relevant sources of information.
- A quantitative evaluation of the presented approach based on realistic mobile device usage data provided by the *Lausanne Data Challenge Campaign* (LDDC) [10].

The remainder of this paper is structured as follows: Section 2 summarizes related work, afterwards Section 3 briefly derives requirements and describes our own approach, that is subsequently evaluated in Section 4. At the end, we summarize our findings, highlight open challenges and give prospects for future work in Section 5.

2. Related Work

As summarized in [15], the task of forecasting the future connectivity of a mobile device can either be seen as a software engineering- or a networking problem. Seen primarily as a software engineering problem, solutions in the domain of mobile cloud computing like *Serendipity* [22] try to distribute computation tasks among other nearby mobile devices to speed up computation or to save energy. Hereby, Serendipity takes into account the future state of the mobile network connection by forecasting its reliability. Similarly, *IC-Cloud* [23] focuses on the challenge of dynamically offloading computation tasks to cloud resources, by taking into consideration that the necessary code and data can be delivered and the results received in time before the next link failure is likely to happen. Specifically developed for the application in mobile edge computing, Sato et al. [20] recently proposed a radio environment aware algorithm that is able to forecast the mobile connectivity between mobile nodes in access points.

Seen primarily as a networking problem, solutions like *BreadCrumbs* [11] try to forecast the future connectivity to WiFi hotspots based on a model of the environment. Recorded sensor data is used to generate user-based models which are then applied to schedule network usage based on connectivity forecasts. BreadCrumbs relies on the fingerprinting of hotspots that is combined with GPS data to forecast a mobile user's bandwidth. Focusing on the aspect of the location forecasting even further, in *NextPlace* [21] a non-linear method is employed to forecast the time and duration of a user's next visit to one of his significant places. Their method identifies patterns in a user's mobility history that are similar to his recent movements in order to forecast his behavior. Similarly, Anagnostopopulus et al. [2] employ supervised learning to perform a classification of trajectories which is then used to forecast the future location of mobile users.

Further related works can either be found in the domain of mobility- and connectivity forecasts like in mobile ad-hoc networking (MANET) or vehicular ad-hoc networking (VANET) [7, 24]. Summarizing the previous findings, it can be concluded that several solutions have been proposed to contribute to the problem of connectivity forecasting. However, current solutions are either not able to operate on a broad range of mobile devices or are not able to provide the level of accuracy, required in IoT scenarios like computation offloading [15].

3. Bandwidth Forecast Service

In this chapter we will first define the design goals for the bandwidth forecast before we choose the dataset to perform our analysis on and decide how to prepare the data. Afterwards, we choose different models that we expect to fit the underlying patterns in the data that are responsible for a changing bandwidth.

3.1. Design Goals

Apart from the functional requirement of a forecast with high accuracy, developing the *Bandwidth Forecast Service* on a mobile device requires to meet several non-functional requirements. Accordingly, we summarize the specific design goals for the development of the Bandwidth Forecast Service with the following key requirements:

- **Support for different time intervals and forecast horizons:** Apart from real-time applications it is often sufficient to forecast the average bandwidth in a certain time interval. The duration of the time interval depends on the use case and can range between minutes and hours. Likewise, the required forecast horizon for the service ranges from the next minute to a bandwidth forecast for the next day.
- **Resource-conserving and customizable:** The service dynamically selects the appropriate learning algorithms based on the required accuracy and the resources available on the mobile device.
- **Can handle small amounts of data:** The service is able to operate even when there is little data available on the mobile device.
- **Privacy aware:** It is possible to process all information on the mobile device itself. The requirement for external data processing should be optional.
- **Open for extension:** The service is open for extension with new learning algorithms or new data sources.

3.2. Dataset and Data Preparation

In January 2009 the Nokia Research Center Lausanne, the Idiap Research Institute and the EPFL initiated the creation of a large-scale mobile data research. This included the design and implementation of the LDDC, an initiative to collect sensor data from smartphones created by almost 200 volunteers in the Lake Geneva region over a period of 18 months [10]. According to [5] it is still the largest dataset that contains information about mobile devices' bandwidth and sensor data, which is why we chose the LDDC dataset to derive the following information as the input to our simulation:

- **Connectivity and bandwidth:** GSM/WiFi/Bluetooth state (on/off), discovered MAC addresses and GSM cells, signal strength of WiFi as well as GSM cells, extended with our own measurements to get an assumption about the available bandwidth.
- **General information about the mobile device itself:** time since the last user interaction, silent mode state, charging state, remaining energy, free memory.
- **Date, time and location:** calendar events, average and estimated remaining duration of stay at the current location.
- **Reasoned attributes:** estimated duration of stay at the same WiFi access point or GSM cell, user is at home/work, travelling, moving, or resting.

This information has been transformed into panel data containing observations over a period of at least 18 months per user. Hereby, we used different time intervals of 2, 10 and 60 minutes to be able to forecast the bandwidth in different granularities. Figure 1 shows the resulting distributions for the GSM bandwidth (left) as well as for the WiFi bandwidth (right).

3.3. Architecture and Model Selection

As a baseline as well as a simple model for mobile devices with very limited resources, we choose an autoregressive model (AR) that tries to forecast the future bandwidth by just taking into account past observations of the forecasting target itself. The *AR(p)* model assumes a linear dependency on its own previous values and a stochastic term. Equation 1 defines the AR model as follows:

$$y_t = c + \sum_{i=1}^{p} \beta_i y_{t-i} + \varepsilon_t \qquad (1)$$

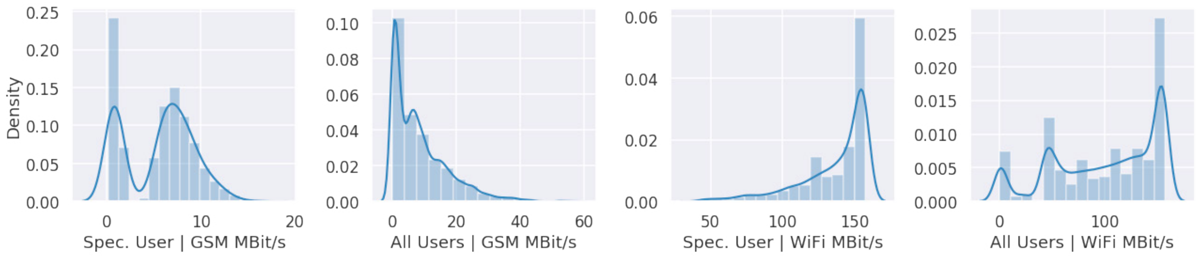

Fig. 1. Distributions of the assumed bandwidth

Based on the AR model, a more sophisticated approach appears to include more information than just the target variable itself. As we assume that arbitrary additional sensor data can be highly multicollinear, we require a type of regression estimator that is well-suited to deal with such types of issues. *Least Absolute Shrinkage and Selection Operator* (LASSO) [25] is an estimator that performs L1 regularization which adds a penalty equal to the absolute value of the magnitude of coefficients encouraging simple, sparse models, i.e. models with fewer parameters. This approach leads to the optimization problem shown in Equation 2, where λ is a nonnegative regularization parameter and p denotes the number of features.

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \frac{1}{N} \sum_{n=1}^{N} (y_n - \beta x_n)^2 + \lambda \sum_{i=1}^{p} |\beta_i| \tag{2}$$

Solving this optimization problem requires substantially more computation power compared to the training of the AR model, but promises to better capture the relationship between the mobile devices' context and its future bandwidth.

Following this approach, we choose decision trees to cover higher order interactions between the individual variables of the sensor data and the target variable. Using gradient boosting [8], which uses ensembles of weak prediction models to iteratively build a stronger model, we aim to build a model that has a low bias and low variance at the same time. As an implementation, we choose *XGBoost* [4] that currently is considered one of the state-of-the-art machine learning algorithms to deal with structured data [12].

3.4. Runtime Environment

We employ standard Java and Android technology to implement the Bandwidth Forecast Service as a microservice which is integrated into the CloudAware mobile middleware presented in [15]. CloudAware is based on the the *Jadex* [18] middleware that provides infrastructure components such as service discovery in mobile environments and allows to expose the microservice.

4. Evaluation

The developed Bandwidth Forecast Service is evaluated by selecting 20 users who have provided data of at least 18 months to the LDDC dataset.

The model used to simulate the usage of a mobile device, described in [16], uses the context data to simulate a mobile device in its constantly changing environment, which aims at reflecting the real-world usage throughout the whole period of the observations. In line with the research goals defined in Section 1, we will focus on two aspects:

- Which context information correlates and helps the most to estimate the future bandwidth of a mobile device?
- Which of the selected models performs best in capturing this relationship?

4.1. Feature Importance

To answer the question which context variable supports the forecast of a future bandwidth at most, we analyze their linear relationship using the Pearson correlation coefficient as shown in Figure 2. Due to the lack of space,

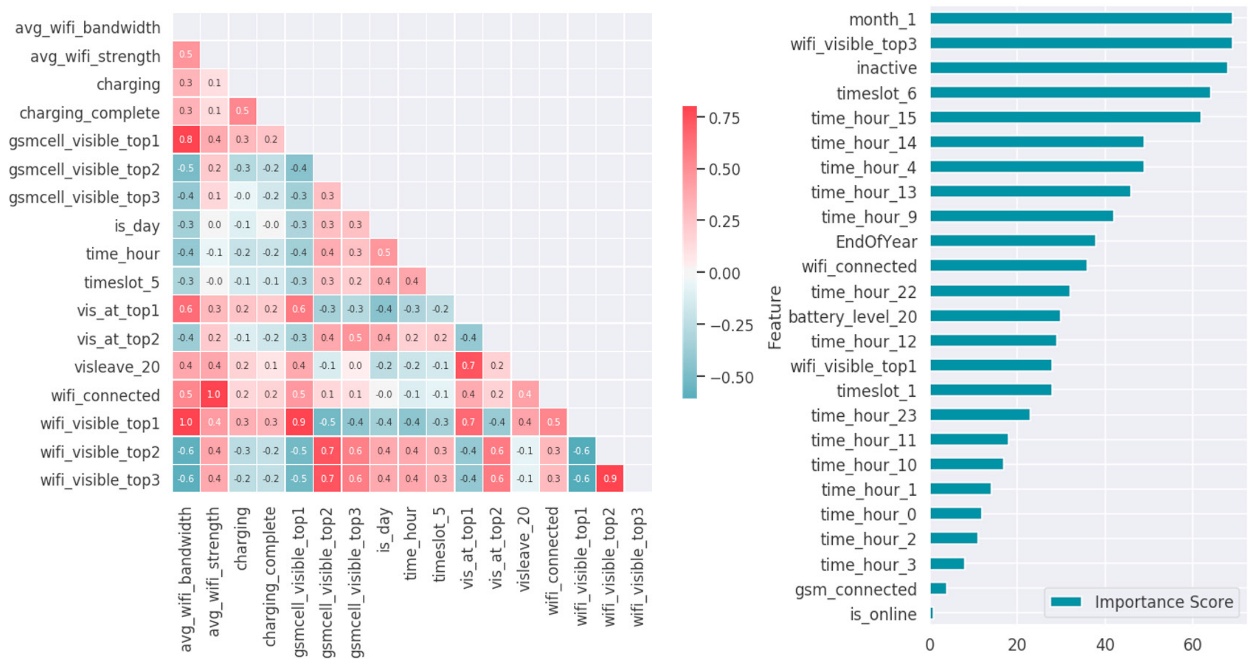

Fig. 2. Correlation matrix for the context variables (left) and feature importance scores of an XGBoost model (right)

we just show the most relevant correlations of the total 112 features contained in the panel. As expected, the current bandwidth, exemplified with the WiFi bandwidth as ”avg_wifi_bandwidth”, is mainly correlated with context variables that describe the current connectivity of the mobile device as well as its physical location.

Although the correlation plot in Figure 2 shows a high correlation between many of the context variables, it supports the hypothesis about the relevance of physical and temporal dependencies. To get a further insight which variables account to the forecast of a future bandwidth at most, we choose the *Feature Importance Score* (F-Score) of the XGBoost model. The F-Score describes the relevance of a feature by counting how many times the underlying tree-based models used the variable as a decision criterion, as shown in Figure 2. Accordingly, it can be concluded that the availability of these variables helps to forecast the future bandwidth of a mobile device. However, this assumption mainly refers to moving devices that users carry with them. Other scenarios in the IoT, apart from the bandwidth forecast, may exhibit other relationships and hence other variables can become the most relevant features for the forecast.

4.2. Simulation Results

The aforementioned simulation is carried out for 20 users of the dataset to validate the general applicability of our approach. Accordingly, we perform an ex-post evaluation using a rolling window approach for which we assume a weekly retraining of the models while, where applicable, the hyperparameters are tuned using grid searches and cross-validation. Although the latter would typically not be carried out in a real-world implementation, it helps to estimate the full potential of the evaluated models. To show the general usefulness of the trained models, Figure 3 presents the actual and the forecasted bandwidth for different context intervals in conjunction with a fixed horizon. Figure 4 shows different forecasting horizons together with a fixed context interval, both on a randomly chosen day of the simulation. A first look at the data leads to the assumption that only the XGBoost model is able to properly capture the relationships between the context data and best forecasts the future bandwidth of a mobile device.

For an overall evaluation of the forecasting accuracy, an appropriate error-measure needs to be selected. As this depends highly on the use-case, we choose the typical linear and quadratic error measures, shown in Table 4.2 for a forecast interval of one hour. Additionally, Figure 5 presents the corresponding error distributions. With a mean absolute percentage error (MAPE) ranging from 35.6% for a bandwidth forecast for the next hour to an MAPE of

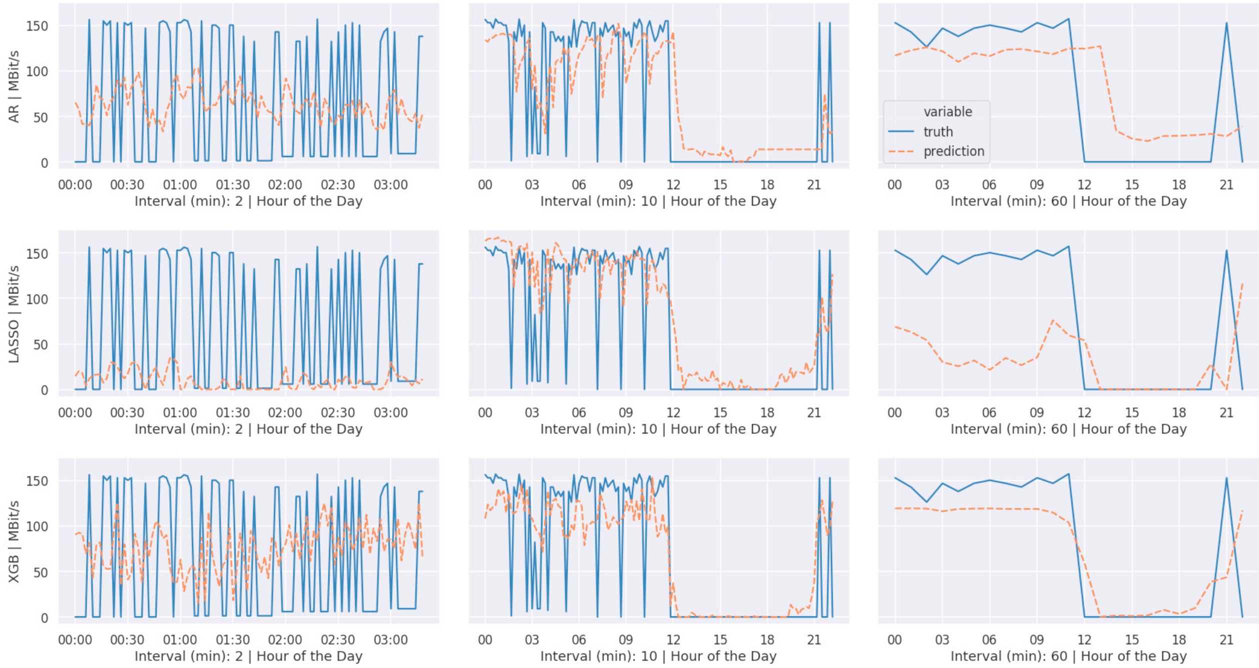

Fig. 3. True and forecasted bandwidth for different models and context intervals, horizon: next interval

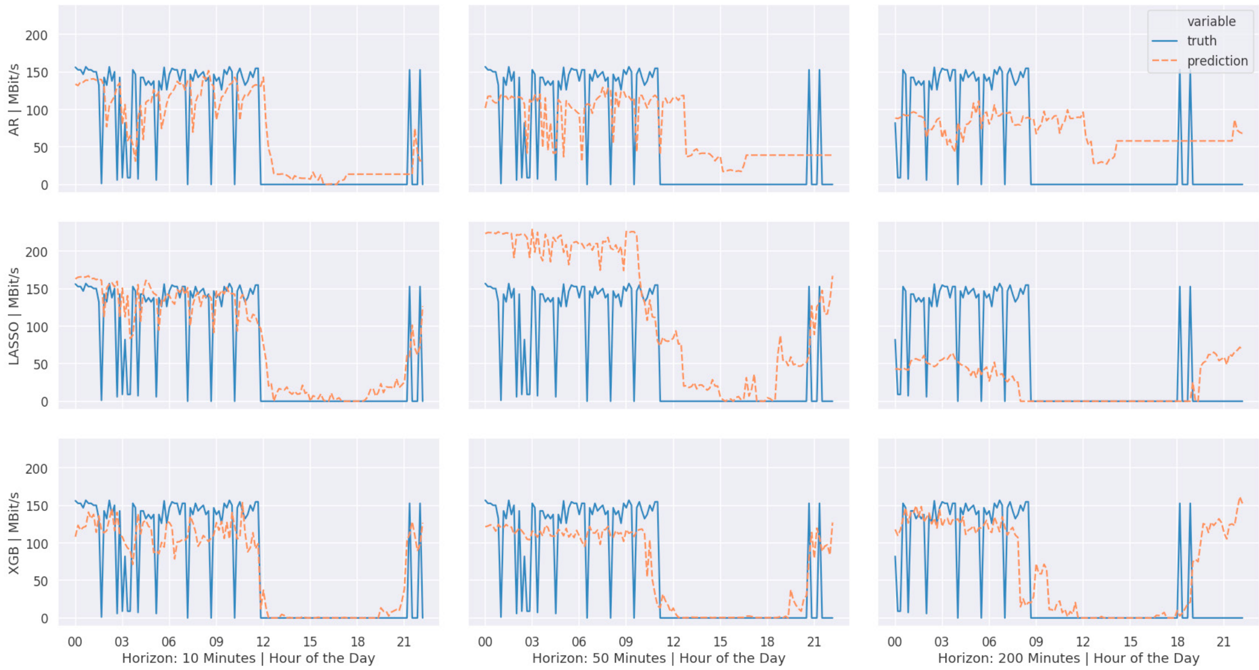

Fig. 4. True and forecasted bandwidth for different models and horizons, interval: 10 minutes

103.3% for a forecast in 20 hours, XGBoost appears to be the most robust and accurate predictor for this problem. Nevertheless, the LASSO and the AR model can depict interesting alternatives, when computation resources are low or the use case does not require such a high level of accuracy.

Table 1. Error measures for different models and horizons

Error	Mean Abs. Error (MAE)			Mean Abs. Perc. Error (MAPE)			Root Mean Squared Error (RMSE)		
Horizon	1	5	20	1	5	20	1	5	20
AR	27,185.3	34,261.9	58.8	38,027.0	51,290.2	103.8	325,301.4	374,053.3	65.6
LASSO	583.2	59.3	66.5	815.8	88.8	117.3	5982.0	82.7	92.0
XGB	25.5	44.2	58.6	35.6	66.2	103.3	36.5	56.1	76.4

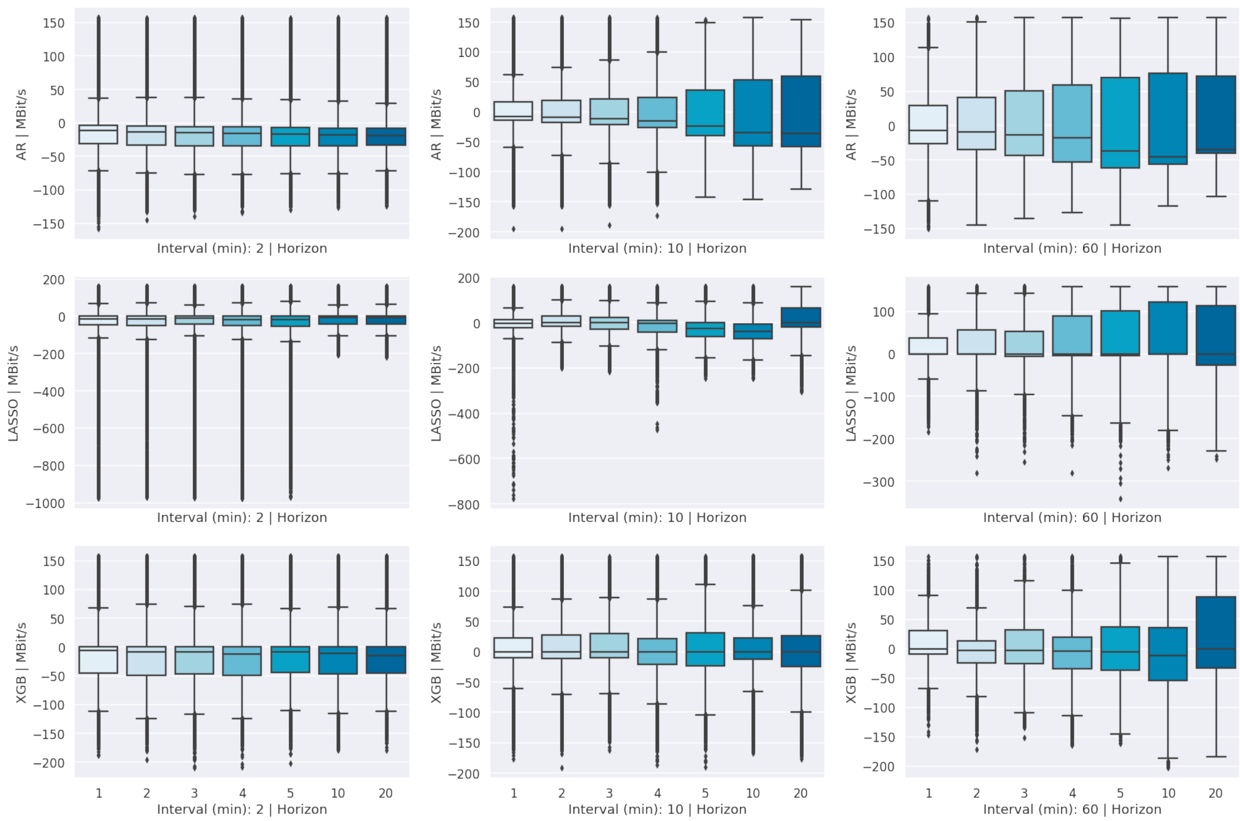

Fig. 5. Distribution of the errors for different models, context intervals and horizons (outliers removed for AR)

5. Conclusion

In this paper, we proposed a novel approach for a Bandwidth Forecast Service for mobile devices in mobile edge and IoT scenarios. Using real usage data of the LDDC we highlighted the benefit of using the mobile devices sensor data to forecast their future bandwidth. We showed that the XGBoost model was able to forecast the future connectivity of mobile devices in a constantly changing environment and we highlighted which sensor data contains the most important information for this task.

This contribution can support a multitude of scenarios where the limited bandwidth of a mobile device needs to be forecasted. Mainly to mitigate the effect of upcoming network bottlenecks by either delaying, advancing or adapting data transfers, but also to save energy. Furthermore, the discovered relationships can also positively influence other scenarios that require to forecast the context of mobile devices.

Nevertheless, reliably forecasting the wireless connectivity by forecasting the user's mobility patterns is considered a complex task and still an open challenge [6, 17]. Although XGBoost was able to provide a high forecast accuracy, its training phase might need to be offloaded to more powerful cloud resources. Hence, at least some categories of mobile devices need to be provisioned with pre-trained models, an aspect we plan to survey in our future work. Especially for the domain of IoT it would also be interesting to further investigate federated approaches, as proposed in [19], that would allow edge devices to form a distributed intelligence.

Acknowledgement: Parts of this research used the MDC Database made available by Idiap Research Institute.

References

[1] Abbas, N., Zhang, Y., Taherkordi, A., Skeie, T., 2018. Mobile edge computing: A survey. IEEE Internet of Things Journal 5, 450–465.
[2] Anagnostopoulos, T., Anagnostopoulos, C., Hadjiefthymiades, S., 2011. Mobility prediction based on machine learning, in: Mobile Data Management (MDM), 2011 12th IEEE International Conference on, IEEE. pp. 27–30.
[3] Apple Inc., 2014. Siri. https://www.apple.com/ios/siri/. Accessed 03.03.2019.
[4] Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA. pp. 785–794. URL: http://doi.acm.org/10.1145/2939672.2939785, doi:10.1145/2939672.2939785.
[5] Crawdad Project, 2019. Crawdad a community resource for archiving wireless data at dartmouth. http://crawdad.org/all-byname.html. Aufgerufen am 03.03.2019.
[6] Farris, I., Taleb, T., Flinck, H., Iera, A., 2018. Providing ultra-short latency to user-centric 5g applications at the mobile network edge. Transactions on Emerging Telecommunications Technologies 29, e3169.
[7] Fernando, N., Loke, S.W., Rahayu, W., 2013. Mobile cloud computing: A survey. Future Generation Computer Systems 29, 84–106.
[8] Friedman, J.H., 2000. Greedy function approximation: A gradient boosting machine. Annals of Statistics 29, 1189–1232.
[9] Johnson, L., . Top 10 location-based mobile advertising campaigns of 2013. http://www.mobilemarketer.com/cms/news/advertising/16821.html. Accessed 03.03.2019.
[10] Laurila, J.K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T.M.T., Dousse, O., Eberle, J., Miettinen, M., 2013. From big smartphone data to worldwide research: The mobile data challenge. Pervasive and Mobile Computing 9, 752–771.
[11] Nicholson, A.J., Noble, B.D., 2008. Breadcrumbs: forecasting mobile connectivity, in: Proceedings of the 14th ACM international conference on Mobile computing and networking, ACM. pp. 46–57.
[12] Nielsen, D., 2016. Tree Boosting With XGBoost - Why Does XGBoost Win "Every" Machine Learning Competition? Dissertation. Norges teknisk-naturvitenskapelige universitet.
[13] Nvidia Corporation, 2015. Nvidia shield game streaming dienst geforce now. http://shield.nvidia.de/game-streaming-with-geforce-now. URL: http://shield.nvidia.de/game-streaming-with-geforce-now. aufgerufen am 03.03.2019.
[14] Orsini, G., Bade, D., Lamersdorf, W., 2016. Generic context adaptation for mobile cloud computing environments, in: The 11th International Conference on Future Networks and Communications (FNC 2016) / The 13th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2016) / Affiliated Workshops, August 15-18, 2016, Montreal, Quebec, Canada, Elsevier Science. pp. 17–24.
[15] Orsini, G., Bade, D., Lamersdorf, W., 2018a. Cloudaware: empowering context-aware self-adaptation for mobile applications. Transactions on Emerging Telecommunications Technologies 29, e3210.
[16] Orsini, G., Bade, D., Lamersdorf, W., 2018b. Generic context adaptation for mobile cloud computing environments (extended version). Journal of Ambient Intelligence and Humanized Computing 9, 61–71.
[17] Patel, P., Ali, M.I., Sheth, A., 2017. On using the intelligent edge for iot analytics. IEEE Intelligent Systems 32, 64–69.
[18] Pokahr, A., Braubach, L., 2013. The active components approach for distributed systems development. International Journal of Parallel, Emergent and Distributed Systems 28, 321–369.
[19] Rahman, H., Rahmani, R., 2018. Enabling distributed intelligence assisted future internet of things controller (fitc). Applied computing and informatics 14, 73–87.
[20] Sato, K., Fujii, T., 2017. Radio environment aware computation offloading with multiple mobile edge computing servers, in: 2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), IEEE. pp. 1–5.
[21] Scellato, S., Musolesi, M., Mascolo, C., Latora, V., Campbell, A.T., 2011. Nextplace: a spatio-temporal prediction framework for pervasive systems, in: International Conference on Pervasive Computing, Springer. pp. 152–169.
[22] Shi, C., Lakafosis, V., Ammar, M.H., Zegura, E.W., 2012. Serendipity: enabling remote computing among intermittently connected mobile devices, in: Proceedings of the thirteenth ACM international symposium on Mobile Ad Hoc Networking and Computing, ACM. pp. 145–154.
[23] Shi, C., Pandurangan, P., Ni, K., Yang, J., Ammar, M., Naik, M., Zegura, E., 2013. IC-Cloud: Computation Offloading to an Intermittently-Connected Cloud. Technical Report GT-CS-13-01. Georgia Institute of Technology.
[24] Shiraz, M., Gani, A., 2014. A lightweight active service migration framework for computational offloading in mobile cloud computing. The Journal of Supercomputing 68, 978–995.
[25] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.