Accepted for presentation in: 2020 IEEE 6th World Forum on Internet of Things (WF-IoT)

Data-Driven Model-Predictive Communication for Resource-Efficient IoT Networks

Christian Arendt, Stefan Böcker, Christian Wietfeld
TU Dortmund University, Germany
Communication Networks Institute (CNI)
Email: {christian.arendt, stefan.boecker, christian.wietfeld}@tu-dortmund.de

***Abstract*—Rapid growth of massive Internet-of-Things (IoT) technologies and a multitude of applications represent an ongoing paradigm shift within our traditional human-centric cellular communication towards expanding connectivity for billions of things. Nowadays, initial generations of connected IoT devices and applications enabled by Cellular-IoT (CIoT) and Low Power Wide Area Networks (LPWAN) technologies are deliberately kept simple and based on equidistant, regular communication intervals. However, this simple communication behavior does not meet 5G requirements of massive Machine Type Communication (mMTC) regarding high node density, long-term energy efficiency and low cost per data. In contrast, this paper presents an approach for an Artificial Intelligence (AI) based model-predictive communication to reduce the effort for continuous, regular data exchange-patterns. Training and test data for the generation of the underlying model are obtained from long-term environment sensor measurements. The derived model approaches are applied to an integrated Industrial-IoT field demonstrator covering a centralized heating control system and decentralized LoRa temperature sensors. Finally, results constitute that overall communication effort per day can be reduced by 60% to more than 95% depending on the required accuracy, significantly contributing to the achievement of mMTC performance targets.**

I. Introduction

The Internet of Things (IoT) is increasingly becoming a reality in people's homes as well as industrial contexts. Emerging long range communication technologies aim to enable a rising amount of sensor applications that have to meet high requirements regarding a high node density, low cost, long battery life and therefore high efficiency in data communications. Requirements defined for 5G massive Machine Type Communication (mMTC) scenarios target a node density of 1.000.000 devices per square kilometer [1]. Two classes of technologies address these requirements using different concepts: Cellular-IoT (CIoT) devices take advantage of exclusively allocated spectral resources in licensed frequency bands in order to enable an interference-free communication. However, the scheduling approach used in licensed frequency bands can only allocate resources for a given amount of devices. Low Power Wide Area Networks (LPWAN), on the other hand, operate in unlicensed frequency bands, in Europe especially the Short Range Device (SRD) bands at 868 MHz and the Industrial Scientific and Medical (ISM) Band at 2.4 GHz. Unlicensed frequency bands can technically be used by anyone and any number of devices. However, regulatory restrictions defined by ETSI [2] apply a fair usage policy of the shared medium, and an interference free channel usage cannot be guaranteed. Typical IoT devices are periodically sending their data in equidistant time intervals. This behavior, while keeping devices simple, does not enable efficient usage of the limited spectral resources available. This work introduces a model-predictive communication framework that allows IoT devices to rate the value of measured sensor data in order to reduce communication effort, leaving spectral resources for other participants. This approach can increase the number of devices being able to get allocated resources in licensed bands as well as reduce the interference probability in unlicensed bands. The concept of this work is illustrated in Fig. 1.

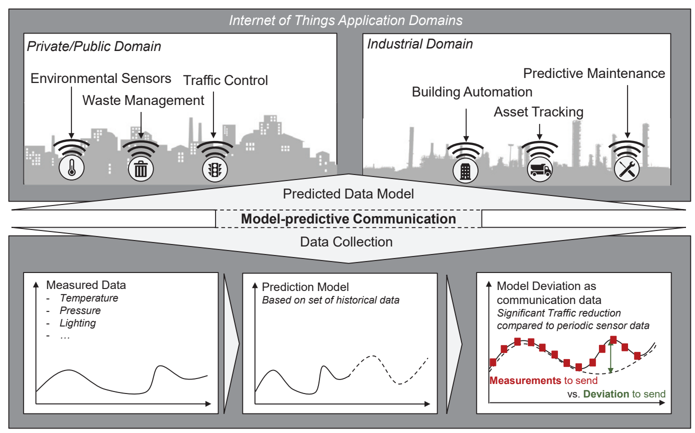


Fig. 1. Model-predictive Communication in Internet of Things Environments

A data-driven approach is proposed in this work, relying on typical data acquisition use cases of sensor networks in public smart city environments as well as Industrial IoT applications like the heating system used as a data source in this work. The application data provided by these sensors is usually measured periodically. To optimize the channel utilization, it is proposed not to transmit every measurement, but only those that show a significant deviation to a predefined model. Therefore, two time series forecasting mechanisms are evaluated in this work to provide such a model of a temperature sensor setting, however, the use case can be adapted easily as these methods only rely on historical data.

The presentation of this work is structured as follows. Section II gives an overview of model-predictive communication frameworks as well as time series prediction models used in communication contexts. The implemented models and the underlying data is depicted in section III. The performance comparison of these models is illustrated in section IV. Transfer of the model in the Industrial-IoT domain is depicted in section V before a conclusion is drawn in section VI.

II. Related Work

As a first step, in [3] we analyzed the capabilites of LoRaWAN networks to contribute to 5G mMTC device density targets. In this work, we provide a method to optimize channel utilization in order to further increase this contribution. The underlying concept of model-predictive communication has been introduced in [4] and revised in [5] for the application of photovoltaic (PV) control data communication. In these works, a deterministic *Efficient PV Production Model* (EPVP) has been developed and used in order to predict the power production of PV systems. However, due to the deterministic nature of the applied model, the adaption to other use cases is not easily possible. Considerable applications and algorithms for Machine Learning in communication networks are presented and discussed in [6], where the authors focus on modeling the network performance. The algorithms of choice for this work, namely Autoregressive Integrated Moving Average (ARIMA) and Long Short Term Memory (LSTM), have been compared for a finance data application in [7]. The authors conclude that LSTM outperforms ARIMA in terms of prediction accuracy provided that enough data is available. The authors of [8] propose an LSTM based analysis of smart meter measurements in order to predict residental load of individual customers in a smart grid system. The LSTM approach is compared to other state of the art machine learning algorithms, such as backpropagation neural networks (BPNN) and k-nearest neighbor (KNN) regression. Presented results conclude that LSTM provides significantly better forecasting results, thus justifying the usage of LSTM in this work. In [9], an Encoder-Decoder LSTM based approach for predicting channel quality in various wireless networks, called *DeepChannel*, is proposed. The authors compare their LSTM approach to an ARIMA approach, concluding that the LSTM based solution performs better when predicting multiple timesteps ahead.

III. Model-Predictive Communication for Internet of Things Applications

As opposed to conventional, periodic data transmissions of IoT sensor applications, this work proposes a data-driven model-predictive approach. This approach uses past sensor data as input for a time series forecasting algorithm for predicting the future data development in the model prediction backend. The predicted series of sensor data can be distributed periodically to the sensor devices and the application management system to determine the deviation of measured sensor data from the prediction. Therefore, only sensor data deviating more than a given tolerance from the prediction model are communicated, resulting in a reduction of communication effort. The deviation is also propagated back to the model prediction backend in order to adjust the input data for the prediction algorithms. A schematic overview of this approach is depicted in figure 2.

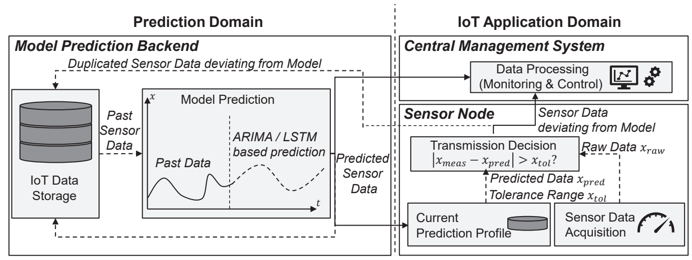


Fig. 2. Schematic Overview of the Model-Predictive Communication approach

In the following section, the underlying dataset as well as the analyzed modeling approaches are introduced.

Model Development based on Private/Public Domain

The dataset used in this work originates from an environmental indoor sensor located in Dortmund, Germany, representing a typical small private or office room environment (private/public domain). The system collected temperature, humidity, and CO_2 concentration with a frequency of roughly 5 minutes from the 1st of January 2019 to 19th of November 2019. In this work, the dataset has been resampled to 30 minutes timesteps to reduce model complexity. The raw data can be accessed via [10].

A. Data-Driven Modeling Approaches

In this work, an autoregression based approach and a neural network approach have been used. Both models leverage the advantages of using a decomposition method, which is described in the following section.

Seasonal and Trend decomposition using Loess (STL)

In this work, Seasonal and Trend decomposition using Loess (STL) method is used for decomposition of an original time series $Y(t) = T(t) + S(t) + R(t)$ into trend $T(t)$, seasonal $S(t)$ and remainder component $R(t)$ [11].

By using this decomposition method, typical properties of the underlying data, such as a daily profile for temperature data, can be extracted. STL is an iterative procedure using mainly two loops. The outer loop calculates robustness coefficients which are used to minimize the impact of outliers, and the inner loop extracts and updates the seasonal and trend components by using Locally Weighted Scatterplot Smoothing (Loess) [12].

The decomposition of sample temperature measurement data used in this work is shown in Fig. 3

The initial data shows a strong daily seasonality profile, as well as a temperature trend corresponding with the weather. The data also shows a significant remainder part which could not be identified by the algorithm as part of the trend or seasonality.

In order to forecast the future development of the time series, forecast algorithms can be used on the deseasonalized time series $Y_{deseas}(t) = T(t) + R(t)$, as the seasonal, periodic component typically changes slowly. Therefore, the extraction

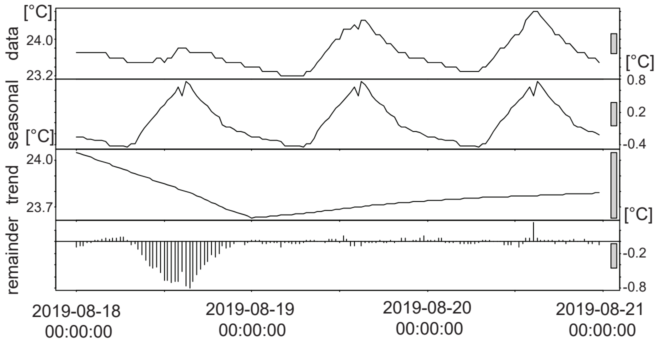


Fig. 3. Decomposition of measured temperature data from 18th of August until 21st of August using STL

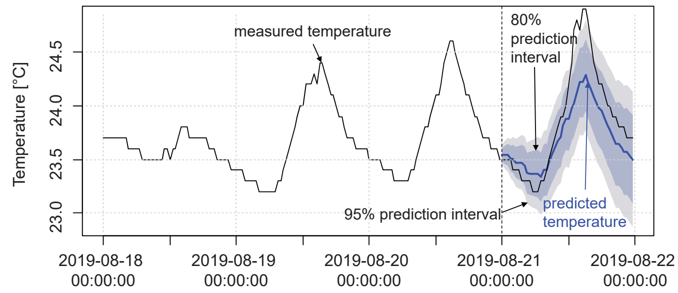


Fig. 4. Forecast for one sensor using STL + ARIMA(0,1,0) for the 21st of August 2019 with 95% and 80% Prediction Intervals

of the seasonal component allows to reduce the complexity of the input data, resulting in a smaller prediction error of the model. The seasonal component is then added back to the forecast time series.

Autoregressive Integrated Moving Average (ARIMA)

One of the most used state of the art algorithms in time series forecasting is the Autoregressive Integrated Moving Average (ARIMA). This algorithm consist of three components, the autoregressive part $AR(p)$, the integrated part $I(d)$ and the moving average part $MA(q)$. $AR(p)$ implies the past values of the series which are used in the prediction model. The parameter p defines how many past values are considered. $I(d)$ refers to the differencing which is often used to make a time series stationary, where d is the degree of differencing. The $MA(q)$ component marks the error of the model in terms of q previous error terms. ARIMA models are typically specified by the parameter set (p, d, q), which defines the parts of the model mentioned above. In this work, an ARIMA implementation for the statistical programming language R is used [13]. This implementation allows the usage of automatic parameter selection for p, d and q for every model realization. d is depicted by a unit test procedure to test for stationarity of the time series, p and q are found by minimizing Akaike's Information Criterion (AIC).

Fig. 4 shows a prediction using ARIMA(0,1,0) for the 2st of August 2019 using a training period of three previous days. In addition to the predicted temperature, the 95% and 80% prediction intervals are shown as well as the actual measurement as test data.

It is shown that ARIMA can achieve a good match between the predicted and the measured temperature for that specific day, resulting in the measured temperature curve lying mostly inside the 80% prediction interval. Figure 5 shows a comparison between the measured and predicted data with different tolerance ranges of 0.5°C, 1°C and 2°C respectively, which means that a deviation of the measured data from the predicted model outside of the chosen tolerance will result in the measured temperature being sent to the heating control system. For this day, a reduction of communication events by 87% can be achieved with a tolerance of 0.5°C. If a deviation below 1°C is sufficient, the maximum reduction potentials are achieved and no communication event would be needed on the given specific day.

Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) networks are a subset of Recurrent Neural Networks (RNN) developed by Hochreiter et al. [14] augmented with the capability to learn long-term dependencies in a dataset. Therefore, they are more suited for time series prediction problems than classic neural network models. LSTM networks consist of concatenated system modules, or cells, connected by constant cell state as well as the data flow through multiple cells. LSTM cells use a "forget gate" to filter past cell states, and an "input gate" to control the influence of the cells input.

This work uses the keras framework on top of the Theano python library for implementing the LSTM model. A network with a single layer is used to keep the training duration short. Additionally, as stated in [9], the usage of deeper networks can even decrease the model accuracy. Within this layer, 50 LSTM cells are utilized, however, the number of LSTM cells did not cause significant variations in model performance.

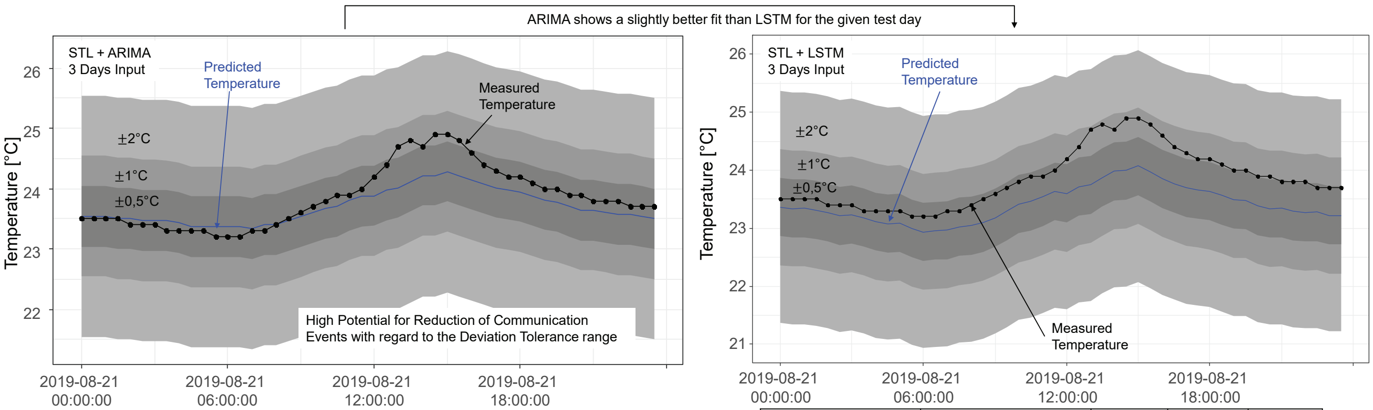
Fig. 5 shows an exemplary prediction for the 21st of August 2019 based on the three previous days in comparison to the ARIMA algorithm.

It can be seen that the LSTM does perform slightly worse than ARIMA. However, depending on the required accuracy, it is still capable of reducing the communication effort by up to 100%.

IV. Statistical Performance Evaluation

In this section both modeling approaches are compared in terms of the modelling *Root Mean Square Error* (RMSE) as well as their potential in reducing communication effort. To draw statistical conclusions, both models have been validated using a sliding window approach depicted in Fig. 6.

The overall dataset is split into sliding sets consisting of a variable amount of consecutive input days followed by one output day. ARIMA is capable of producing a sufficient prediction solely based on small amounts of data. In this work, we analyzed impact of varying input days from 3 to 12, to predict one forecast day. In contrast, LSTM relies on large amounts of training data, splitted into multiple input/output



Tolerance range	Regular 30 Min Interval	±0.5°C	±1°C	±2°C
Communication events	48 (100%)	6 (13%)	0 (0%)	0 (0%)

Tolerance range	Regular 30 Min Interval	±0.5°C	±1°C	±2°C
Communication events	48 (100%)	20 (42%)	0 (%)	0 (0%)

Fig. 5. Forecast for one sensor using STL + ARIMA(0,1,0) and LSTM for the 21nd of August 2019 with tolerance ranges of ± 0.5°C,± 1°C and ± 2°C. ARIMA shows a slightly higher potential in decreasing communication effort for ± 0.5°C tolerance.

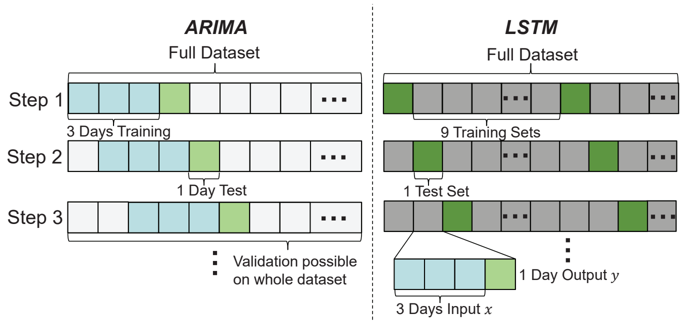


Fig. 6. Walk Forward Validation with Sliding Window (Constant Training Period Length) in each algorithm's variant. ARIMA takes t days training data and produces a one day forecast which is compared to the test day, LSTM needs input/output pairs for training and is then tested on a held back test set. A 10-fold cross validation approach is used to enable validation on the whole dataset.

pairs, to learn necessary features that enable the prediction of forecast sensor series. To further verify the forecast result, a 10-fold cross-validation approach, with a split of 90-to-10% between training and test data, is applied.

Fig. 7 illustrates the RMSE distributions of both approaches for varying input periods.

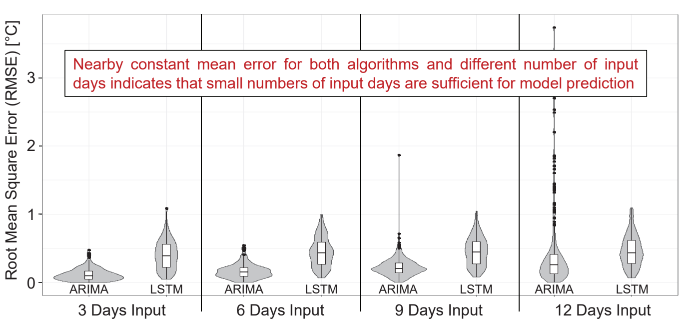


Fig. 7. Root Mean Square Error (RMSE) for both approaches and varying training period. ARIMA shows a smaller RMSE than LSTM, however the mean error for both approaches remains nearly constant.

Both models exhibit a smaller error with smaller training periods, however, the training period length does not show a significant impact on the accuracy of both models. ARIMA shows a smaller mean error for all training periods (around 0.15°C) than LSTM (around 0.3°C). Except a small amount of outliers, ARIMA indicates a smaller error spread. The error distributions of the models indicate an estimation for sufficient tolerance ranges in which the models can be applied. For the analysis of the impact the model-predictive approach has on communication effort, we therefore defined ± 0.5°C as the smallest tolerance range for the model. As a sensitivity analysis, two additional tolerance ranges have been defined, namely ± 1°C and ± 2°C. Fig. 8 then depicts the potential in reducing communication effort for the underlying temperature sensor system. Both models exhibit a potential of more than 60% reduction of communication effort even with a tolerance of ± 0.5°C. Even with our minimum defined tolerance range of ± 0.5°C, both models are able to reduce communication effort by more than 60% for LSTM and about 80% for ARIMA. While Arima only slightly profits from an increased number of input days, LTSM performance is reduced for longer input periods, as these result in an overall smaller amount of datasets used for training. For higher tolerance ranges, both models perform nearly equally well, resulting in a communication effort reduced by more than 90% (± 1°C) or nearly no communication effort (± 2°C), respectively. In conclusion, both model approaches exhibit a significant potential to reduce communication effort of sensor systems. The ARIMA approach is a preferred solution as it has an overall better performance and is a much more efficient model in terms of computational effort and input data requirements.

V. Application on Industrial-IoT Domain

In order to proof applicability of our developed model approaches in the Industrial-IoT domain, we have evaluated the optimization potential for reducing communication effort to a dataset derived from a large industrial production site. As depicted in Fig. 9, the manufacturing environment represents a

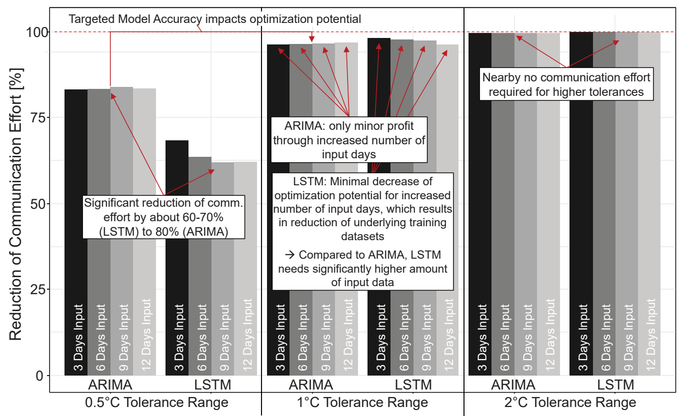


Fig. 8. ARIMA shows an relatively constant high potential in communication reduction with increasing input period, while LSTM performs worse, especially with longer input periods, as these result in smaller training datasets. Depending on the desired accuracy, Communication effort can be reduced by 60% to nearly 100%

more demanding heating use case. The system environment is divided into two parts. First, intelligent temperature sensors are installed within a large production environment (manufacturing hall, ground floor) to update the temperature information at a regular interval of 30 minutes. In total 4 sensors are installed, which communicate the measured sensor information via a LoRa Peer-to-Peer (P2P) link to a central heating system, located in the basement of the production site. The heating system, consisting of the physical pump and the intelligent heating control unit, processes the individual input data of each sensor and determines an actuating variable for the heating control.

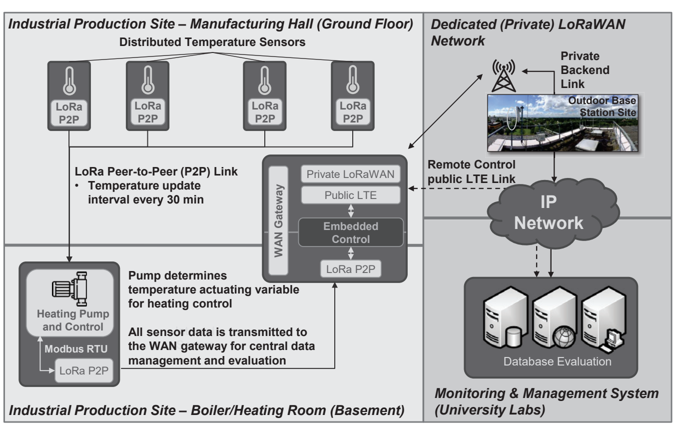


Fig. 9. System overview of real field environment in industrial production environment. In our scenario, the central heating control is supported through the utilization of interconnected temperature sensors.

On the other side, the heating pump and control unit forwards single sensor information as well as determined actuating variable to the Wide Area Network (WAN) gateway. The WAN gateway serves as interface between internal production site and a central management system. WAN access is established over a dedicated LoRaWAN network and is utilized as a low-cost private back-end data link. Additionally, the WAN Gateway is equipped with an integrated LTE modem for remote control and maintenance purposes. In the following, the centrally aggregated sensor data from 29th of June until the 29th of October 2019 is analyzed and processed with regard to optimized ressource-efficient communication behaviour.

The raw input data received from the setup described in this section has to be preprocessed in order to be used for the forecasting mechanisms described in the previous sections. Due to imperfect timing of the sensor nodes, the measured datapoints are not exactly equally spaced. Additionally, some datapoints are missing because of packet losses. Within the evaluation of the setup, there have been power outages in the industrial site, leading to a loss of data for longer time frames. In order to apply the forecasting algorithms, a resampling approach has been used together with an interpolation, taking the seasonality of the dataset into account. In this case a 4-fold cross-validation is applied, due to the reason that the overall dataset is limited to a period of about 4 months. Fig. 10 shows the communication reduction potential in this scenario.

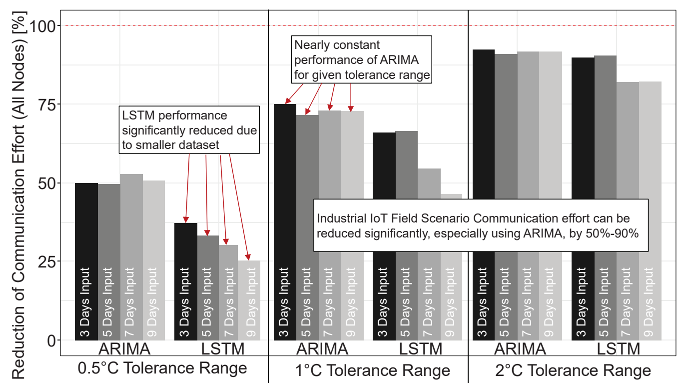


Fig. 10. Communication effort in the Industrial-IoT field scenario can be reduced significantly.

It can be seen that the characteristic between both models is the same as in the private/public domain and, even in this more demanding environment, a communication effort reduction of 30% to 80% (LSTM) or 50% to 80% (ARIMA) can be achieved depending on the required accuracy, proving the applicability of the modeling approaches for both the public/private domain and the Industrial-IoT domain.

VI. Conclusion

This work presents a model-predictive communication framework based on the analysis of historical data. The data used is derived from an indoor environmental sensor system and the models have been adapted to a real world Industrial-IoT heating control environment, which defaults to periodical transmissions every 30 minutes. To reduce this communication effort, two prediction algorithms have been implemented and compared, namely the autoregression based ARIMA and a neural network with LSTM cells. The analysis shows a significant potential to reduce communication effort using these prediction schemes of at least 60% up to more than 90% depending on the required model accuracy. It can be seen that ARIMA shows a better performance in this scenario. Together

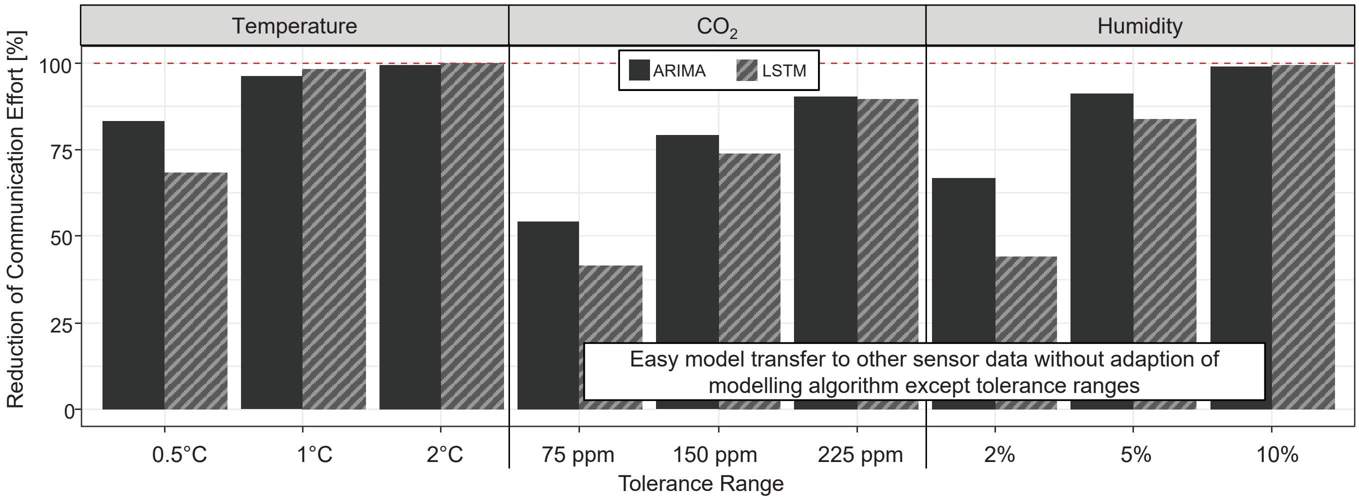


Fig. 11. Modeling approaches in this work can easily be adapted to different sensor data scenarios and tolerance ranges

with the significantly smaller computation effort needed by the ARIMA model, this model is able to substantially reduce the resource demand of IoT systems in terms of spectral resources.

The developed model approach based on temperature data can be easily transferred to various sensor datasets. This is confirmed by a sensitivity analysis applied on humidity and CO_2 concentration sensor data with 3 input days, produced by the same environmental sensor as the initial temperature dataset (see Fig. 11). It can be seen that, even for stricter tolerances for CO_2 concentration and humidity, our approach allows a substantially reduced communication effort. Therefore, the proposed data driven model-predictive communication scheme can significantly increase the contribution of various IoT-technologies to the 5G mMTC targets of one million devices per square kilometre by reducing the communication demand per device. Especially LPWAN Technologies like LoRaWAN can therefore have a substantially higher impact on mMTC target fulfillment [3].

In future work, the Industrial-IoT heating system setup will be enhanced by the introduced model-predictive framework for evaluation. In order to further improve model accuracy, external factors such as the outdoor temperature and weather forecast, will be taken into account.

Acknowledgment

Part of the work on this paper has been supported by the federal state of Northrhine-Westphalia and the "European Regional Development Fund" (EFRE) 2014 - 2020 in the course of the "OPUS" project under grant number EFRE-0800885, as well as the German Federal Ministry of Education and Research (BMBF) in the project A-DRZ (13N14857) and the Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Analysis", project A4.

References

[1] ITU, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," International Telecommunication Union, Tech. Rep. ITU-R M.2410-0, 2017.

[2] European Telecommunications Standards Institute, "Short range devices (SRD) operating in the frequency range 25 MHz to 1 000 MHz; part 2: Harmonised standard for access to radio spectrum for non specific radio equipment," European Telecommunications Standards Institute, Tech. Rep., 2018.

[3] S. Böcker, C. Arendt, P. Jörke, and C. Wietfeld, "LPWAN in the context of 5G: Capability of LoRaWAN to contribute to mMTC," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, Limerick, Ireland, Apr. 2019, pp. 737–742.

[4] K. R. Piontek, N. Dorsch, and C. Wietfeld, "Resource-efficient model-predictive PV control data communication," in *Innovative Smart Grid Technologies - Asia (ISGT Asia), 2014 IEEE*, May 2014, pp. 715–720.

[5] T. Kleinschmidt and C. Wietfeld, "Fluctuation-sensitive model-predictive communication for distributed energy resources," in *2016 IEEE International Conference on Communications Workshops, ICC 2016*, 2016.

[6] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 3, pp. 1790–1821, 2017.

[7] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A comparison of ARIMA and LSTM in forecasting time series," *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, pp. 1394–1401, 2019.

[8] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[9] A. Kulkarni, A. Seetharam, A. Ramesh, and J. Herath, "DeepChannel: Wireless channel quality prediction using deep learning," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2019.

[10] C. Arendt, "Raw environmental indoor sensor data," Nov 2019. [Online]. Available: http://doi.org/10.5281/zenodo.3555406

[11] Robert B Cleveland, William S. Cleveland, Jean E. McRae, and Irma Terpenning, "STL: A seasonal-trend decomposition procedure based on loess," pp. 3–73, 1990. [Online]. Available: http://www.nniiem.ru/file/news/2016/stl-statistical-model.pdf

[12] W. S. Cleveland and S. J. Devlin, "Locally weighted regression : An approach to regression analysis by local fitting locally weighted regression : An approach to regression analysis by local fiiting," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 37–41, 1988.

[13] R. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for r," *Journal of Statistical Software, Articles*, vol. 27, no. 3, pp. 1–22, 2008. [Online]. Available: https://www.jstatsoft.org/v027/i03

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.