A Reinforcement Learning Approach for Efficient Opportunistic Vehicle-to-Cloud Data Transfer

Benjamin Sliwa and Christian Wietfeld

Communication Networks Institute, TU Dortmund University, 44227 Dortmund, Germany
e-mail: {Benjamin.Sliwa, Christian.Wietfeld}@tu-dortmund.de

Abstract—**Vehicular crowdsensing is anticipated to become a key catalyst for data-driven optimization in the Intelligent Transportation System (ITS) domain. Yet, the expected growth in massive Machine-type Communication (mMTC) caused by vehicle-to-cloud transmissions will confront the cellular network infrastructure with great capacity-related challenges. A cognitive way for achieving relief without introducing additional physical infrastructure is the application of opportunistic data transfer for delay-tolerant applications. Hereby, the clients schedule their data transmissions in a channel-aware manner in order to avoid retransmissions and interference with other cell users. In this paper, we introduce a novel approach for this type of resource-aware data transfer which brings together supervised learning for network quality prediction with reinforcement learning-based decision making. The performance evaluation is carried out using data-driven network simulation and real world experiments in the public cellular networks of multiple Mobile Network Operators (MNOs) in different scenarios. The proposed transmission scheme significantly outperforms state-of-the-art probabilistic approaches in most scenarios and achieves data rate improvements of up to 181% in uplink and up to 270% in downlink transmission direction in comparison to conventional periodic data transfer.**

I. INTRODUCTION

Within the ongoing transition from human-controlled cars to autonomous traffic systems, the exploitation of the vehicles themselves as moving sensor nodes is a key enabler for data-driven traffic optimization. However, while future ITSs will significantly benefit from high penetration rates of environment-sensing vehicles, the cellular network infrastructure will be confronted with massive increases in resource occupation related to Machine-to-machine (M2M) communication. A promising approach for improving the resource efficiency of the *existing* network infrastructure is to apply *opportunistic* communication techniques for data-intense *delay-tolerant* applications. In order to avoid wasting cell resources on packet error-related retransmissions, data transmissions are performed in a channel-aware manner with respect to the expected network quality with the goal of minimizing the error probability.

Similarly, the *anticipatory* communication paradigm [1] proposes to explicitly consider context information (e.g., measurements of the network quality) for optimizing decision processes within communication networks. In previous work, we have presented the client-based opportunistic transmission schemes Channel-aware Transmission (CAT), predictive CAT (pCAT), and Machine Learning CAT (ML-CAT) [2], [3], which schedule vehicular sensor data transmissions with respect to the expected network quality. Although those methods

are able to achieve massive improvements in the resulting end-to-end data rate and power efficiency, they are a based on heuristics and their large parameter spaces complicate the determination of the optimal operating point.

In this paper, we extend the general ideas of context-aware opportunistic data transfer with a hybrid cognitive networking approach, which brings together *reinforcement learning*-based decision making with supervised machine learning for data rate prediction. The mobile device is modeled as a context-sensing agent, which autonomously learns to detect and exploit favorable transmission opportunities on its own by only considering previously taken actions. For this form of *exploration*, we exploit the findings and the open data set of a large-scale real world data rate prediction campaign, which was presented in [4]. Fig. 1 illustrates the real world channel dynamics by means of an Signal-to-noise-plus-interference Ratio (SINR) time series trace and illustrates the involved challenges and opportunities for the vehicular data transfer. Instead of only using SINR measurements for channel quality assessment, we jointly consider nine different Long Term Evolution (LTE) network quality indicators within the proposed reinforcement learning-based approach, which are brought together by a data rate prediction model.

The contributions provided by this paper are as follows:

- Presentation of the novel **reinforcement learning-based** transmission schemes Reinforcement Learning CAT (RL-CAT) and Reinforcement Learning pCAT (RL-pCAT) for optimizing the resource efficiency of cellular vehicular sensor data transmissions.
- Proof-of-concept **real world performance evaluation** and comparison with state-of-the-art probabilistic data transfer approaches.
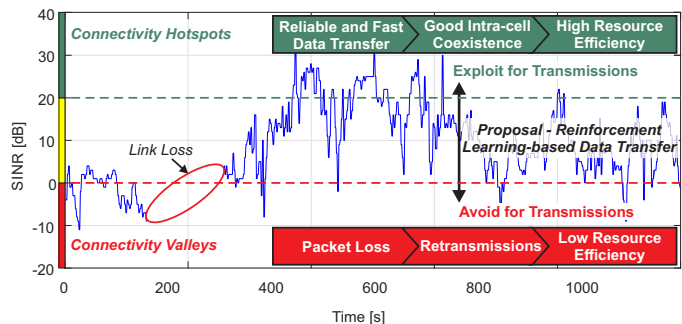

Fig. 1. Interdependency between network quality and end-to-end transmission behavior. In this paper, reinforcement learning is applied to exploit *connectivity hotspots* and to avoid *connectivity valleys* for reaching the overall goal of improving the resource efficiency of the network.

- The measurement and evaluation software[1] as well as the raw results of the performance evaluation [5] are provided in an **open source** way.

The remainder of the paper is structured as follows. In Sec. II, we give an overview about data-driven optimization approaches for vehicular data transfer. Afterwards, the proposed reinforcement learning-based transmission schemes are presented in Sec. III. An overview about the methodological evaluation setup is provided in Sec. IV. Finally, the results of the performance evaluation are presented and discussed in Sec. V.

II. RELATED WORK

Machine learning is a catalyst for anticipatory communication in complex environments, as it allows to implicitly consider the hidden interdependency between observable measurement variables, which are too complex to bring together in a closed analytical description. An overview about different machine learning approaches for wireless communication systems is given in [6]. Furthermore, the authors of [7] provide a comprehensive overview about machine learning methods for vehicular communication networks. *Supervised* learning models utilize *labeled* data for training of *regression* models, which can then be used for performing *predictions* on unlabeled data. Typically, the models are trained offline and then deployed to the target platform for online application. If major changes of the environment occur – e.g., a previously unobserved packet scheduler is deployed to the cellular network by the MNO – novel data needs to be obtained and the prediction models need to be re-trained.

In contrast to that, *reinforcement learning* [8] introduces the concept of *cognitive* decision making, where a virtual *agent* senses the environment and optimizes a certain behavior by learning from the *rewards* of taken *actions*. A comprehensive overview about applying this type of machine learning to mobile communication systems is provided by [9].

Data rate prediction has been proposed as a method for increasing the context-awareness of vehicular communication systems through high-level optimization techniques, e.g., for predictive caching [10] and multi-Radio Access Technology (RAT) interface selection [11]. Consequently, different research works have investigated client-based data rate prediction in vehicular cellular networks, for which the main findings are summarized as follows:

- The radio channel characteristics have a severe impact on the resulting end-to-end data rate. Passively measurable indicators such as Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ) and SINR can provide meaningful information for data rate prediction models [3], [4], [12], [13], [14].
- Integrating the payload size into the prediction process allows to implicitly consider cross layer interdependencies (e.g., the slow start mechanism of Transmission Control Protocol (TCP)), which have a strong interdependency with the channel coherence time [3], [4].

---

[1]Source code is available at https://github.com/BenSliwa/MTCApp

- Crowdsensed connectivity maps can be applied to maintain radio condition data bases, which allow to forecast the network situations vehicles are going to encounter on their future trajectories. Furthermore, the applied cell-wise aggregation implicitly compensates short-term prediction errors (e.g., related to multipath fading) [3], [4].
- In the vast majority of the evaluations (e.g., [3], [4], [14], [15]), Classification and Regression Tree (CART)-based methods such as Random Forests (RFs) [16] outperform more complex models (e.g., *deep learning* [17]), which require a significantly higher amount of training data.
- Cellular data rate prediction models are highly MNOs-dependent due to MNOs-specific configurations of the network infrastructure [4].

Data-driven Network Simulation (DDNS): In addition to using data rate prediction models for context-aware decision making, the trained models themselves can provide the foundation for simulative optimization of cognitive communication systems using DDNS [18]. This approach allows to replay available time series measurements of *passive* network quality indicators to analyze the behavior of novel *active* transmission schemes. The end-to-end behavior is represented by a supervised machine learning model, for which the deviation to the real world measurements is learned by a second Gaussian Process Regression (GPR)-based machine learning model. The latter transfers the prediction process from the deterministic to the probabilistic domain and allows to generate synthetic – yet close to reality – end-to-end indicator profiles by sampling from distribution of the prediction errors. As the real world validation in [18] shows, the achieved results are not only significantly more accurate than conventional system-level network simulations [19], the result generation process is also more than an order of magnitude faster, which ultimately allows to perform a deeper exploration of the parameter space within the system optimization phase.

Opportunistic data transfer is a method for optimizing the resource efficiency of data transmissions for *delay-tolerant* applications by integrating the network quality into the transmission process. Recently, the authors of [20] have presented Client-side Adaptive Scheduler That minimizes Load and Energy (CASTLE) as a method for distributed client-side scheduling of coordinated transmissions, which exploits machine learning for channel-sensitive load estimation. In previous work [2], [3], we have applied probabilistic methods, where a transmissions probability $p_{\text{TX}}(t)$ is calculated based on measurements of a network quality indicator $\Phi(t)$ with a defined value range $\Phi_{\max} - \Phi_{\min}$, application-specific deadlines $\Delta t_{\min}$ and $\Delta t_{\max}$ and a convergence exponent α as

$$p_{\text{TX}}(t) = \begin{cases} 0 & \Delta t < \Delta t_{\min} \\ 1 & \Delta t > \Delta t_{\max} \\ \left(\frac{\Phi(t) - \Phi_{\min}}{\Phi_{\max} - \Phi_{\min}}\right)^{\alpha} & \text{else} \end{cases} \quad (1)$$

where Δt is the passed time since the last successful transmission has been performed. For the basic transmission scheme
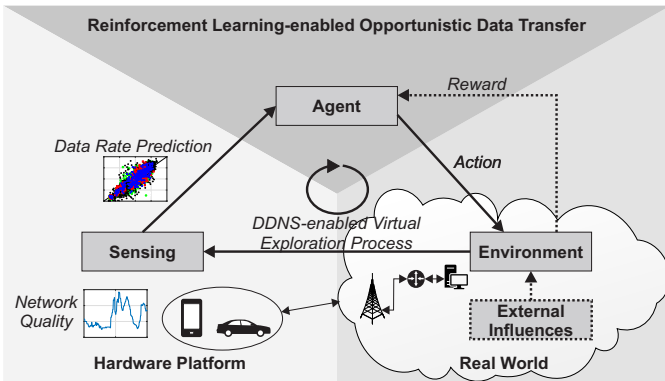
Fig. 2. System architecture model: Interaction between agent and environment.

CAT, the transmission metric Φ is represented by the measured SINR, while ML-CAT considers the predicted data rate. In this work, the goal is to further optimize the data rate improvement by replacing the probabilistic approach with reinforcement learning-based decision making.

III. REINFORCEMENT LEARNING-BASED OPPORTUNISTIC VEHICULAR DATA TRANSFER

In this section, we present the proposed reinforcement learning-based transmission schemes. The overall goal is to learn a context-aware transmission process, which *exploits* connectivity hotspots and *avoids* data transmissions during connectivity valleys. The overall system architecture model and the interaction between *agent* – the opportunistic transmission scheme – and *environment* – the public cellular network – is illustrated in Fig. 2. The proposed system is composed of three logical domains. The actual decision making is performed in the *agent* domain, where the agents decides if the buffered data shall be transmitted immediately or if the transmission should be postponed as the current network situation is not favorable. The *real world* domain represents the network environment, which mainly impacted by external influence factors (e.g., other cell users, mobility-related channel dynamics) and not by the taken actions of the agent itself. For completeness, it is remarked that transmissions performed by the agent have a minor impact on the environment due to the occupied network resources. The foundation for the decision making is the *sensing* process, which is performed in the *hardware platform* domain. Based on measurements of raw context features, a prediction model is applied to forecast the currently achievable data rate. In this paper, we exploit the high computational efficiency of DDNS for implementing a *virtual* exploration process (see Sec. IV-B), which trains the agent with data transmission profiles synthesized from previous real world transmissions. Based on this foundation, we derive the reinforcement learning-based opportunistic data transmission scheme RL-CAT and its mobility-predictive extension RL-pCAT in the next paragraphs.

A. Context-aware Approach: Reinforcement Learning CAT (RL-CAT)

Reinforcement learning is applied to derive a *decision table* Q, which allows to assess the expected rewards by performing the possible actions for a given state. For opportunistic data transfer, the possible actions a are IDLE (data is buffered) and TX (data is transmitted). We model the state as a *context tuple* $\mathbf{C}_t$ as

$$\mathbf{C}_t = (\tilde{S}(t), \Delta t) \qquad (2)$$

with $\tilde{S}(t)$ being the predicted data rate, which is discretized to the closest integer value, and Δt being the passed time since the last successful transmission has been performed. As the data rate prediction accuracy is reduced in the edge regions of the data rate value range [4], the reinforcement learning process immanently learns a confidence model for the machine learning-based data rate prediction.

Online Decision Making: An overview of the process for online decision making is illustrated in Fig. 3. Different context measurements are used to estimate the currently achievable data rate $\tilde{S}(t)$ with a supervised prediction model. Based on the findings and analyzed model comparisons in [4], the latter is represented by a RF prediction model with maximum depth 15 and 100 trees. The feature vector $\mathbf{F}(t)$ is composed of measurements from different context domains:

- **Network**: RSRP, RSRQ, SINR, Channel Quality Indicator (CQI), Timing Advance (TA), Carrier frequency f
- **Mobility**: Velocity, Cell id
- **Application**: Payload size of the to be transmitted data packet

Using the predicted data rate $\tilde{S}(t)$ within $\mathbf{C_t}$ instead of the raw context features significantly reduces the dimension of the table index. Therefore, the exploration phase of the reinforcement learning process can be performed more efficiently, as fewer exploration epochs are required. Together with Δt, the context state $\mathbf{C_t}$ is composed and the to be performed action a is then selected by maximizing the achievable Q value with $\max_a Q(\mathbf{C_t}, a)$.
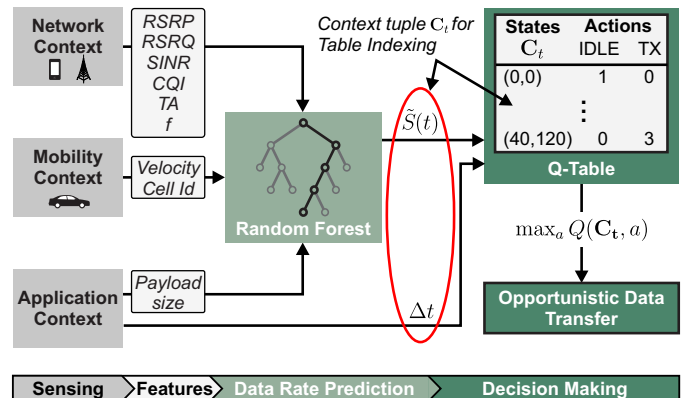

Fig. 3. Online application of the reinforcement learning-based decision making for the basic RL-CAT approach. For RL-pCAT, the context state is represented by an extended tuple $\mathbf{C}'_t = (\tilde{S}(t), \tilde{S}(t+\tau), \Delta t)$.

Iterative Exploration Process: Before the system is able to make cognitive decisions on its own, it needs to fill the Q-table with valid data through an iterative *exploration* process, which considers the rewards of previously performed actions. For the proposed reinforcement learning-based transmission scheme, we apply an adjusted version of the classical *Q-learning* [21] technique. At first, all table entries are randomized with $\mathcal{N}(0,1)$. For each performed action, the Q-table is then updated as

$$Q(\mathbf{C_t}, a) = (1-\alpha) \cdot Q(\mathbf{C_t}, a) + \alpha \left[r_a + \lambda \cdot \max_a Q(\mathbf{C_{t+1}}, a) \right] \tag{3}$$

with *learning rate* α, *reward* r_a, *discount factor* λ, and $\mathbf{C_{t+1}}$ being an estimation of the Q-value in the future state after the action a has been taken. However, in the considered vehicular scenario, the environment state is changed mainly due to external impact factors – e.g., mobility-related channel dynamics – and the taken actions do not have a measurable impact. Therefore, Eq. 3 is simplified to

$$Q(\mathbf{C_t}, a) = (1-\alpha) \cdot Q(\mathbf{C_t}, a) + \alpha \cdot r_a \tag{4}$$

Separate reward functions r_a are applied for the possible actions a. The reward r_{TX} of a performed transmission with measured data rate S is calculated with respect to the trade-off between a defined MNO-specific target data rate S^* and an application-specific age of information deadline Δt_{max} with a weighting trade-off factor w

$$r_{\text{TX}}(S, \Delta t) = \frac{w \cdot (S - S^*)}{S_{\max}} + \frac{\Delta t \cdot (1 - w)}{\Delta t_{\max}} \tag{5}$$

For RL-CAT, the `IDLE` action is not able to achieve a reward as no data is transferred during the buffering phase. However, a *deadline violation punishment* which is represented by a large negative number Ω is introduced in order to ensure $Q(\mathbf{C_t}, \text{TX}) \gg Q(\mathbf{C_t}, \text{IDLE})$ if the Age of Information (AoI) deadline is reached, which then causes an immediate data transmission regardless of the expected resource efficiency.

$$r_{\text{IDLE}}(\Delta t) = \begin{cases} \Omega & \Delta t \geq \Delta t_{\max} \\ 0 & \text{else} \end{cases} \tag{6}$$

B. Context-predictive Approach: Reinforcement Learning pCAT (RL-pCAT)

As previous studies [3] have shown, opportunistic vehicular data transfer methods can significantly benefit from not only considering the current context, but also taking predictions for the anticipated future context behavior into account. In the following, we therefore extend the basic RL-CAT concepts to the context-predictive RL-pCAT method which considers the anticipated future network quality along the expected trajectory of the vehicle. The context tuple is extended by an additional data rate prediction $\tilde{S}(t + \tau)$ for a given temporal look ahead τ to $\mathbf{C}'_t$ as

$$\mathbf{C}'_t = (\tilde{S}(t), \tilde{S}(t + \tau), \Delta t) \tag{7}$$

However, as the future feature vector $\mathbf{F}(t + \tau)$ cannot be measured at the time of the decision making t, it is predicted

based on aggregated measurements which were previously performed in the same geographical region. The estimated feature vector $\tilde{\mathbf{F}}(t + \tau)$ is looked up from a multidimensional connectivity map M with cell size c as

$$\tilde{\mathbf{F}}(t + \tau) = M(\lfloor \frac{\tilde{\mathbf{P}}(t + \tau)}{c} \rfloor) \tag{8}$$

with $\tilde{\mathbf{P}}(t + \tau)$ being an estimation of the future vehicle position, which is derived based on trajectory-aware mobility prediction. Details about the algorithmic implementation and a real world evaluation of the prediction errors as well as their impact on the network quality prediction can be found in [22].

The reinforcement learning process is performed analogously to Sec. III-A. However, the reward function of the `IDLE` action is changed to r'_{IDLE} as

$$r'_{\text{IDLE}}(\Delta t) = \begin{cases} \Omega & \Delta t \geq \Delta t_{\max} \\ \frac{1}{\tau} \cdot r_{\text{TX}}(\tilde{S}(t + \tau), \Delta t + \tau) & \text{else} \end{cases} \tag{9}$$

as postponing the transmission at t is now immanently related to the predicted context at $t + \tau$.

It is remarked that the application of context prediction introduces additional error sources to the system which impact the achievable performance. Imperfections of the mobility prediction mechanism might lead to false context lookups and the context aggregation within the connectivity map only represents the mean indicator behavior within the considered cell. In addition, as the data rate prediction is performed for $\tilde{S}(t)$ as well as for $\tilde{S}(t+\tau)$, prediction errors have an increased impact on the channel quality assessment.

IV. METHODOLOGY

In this section, the methods for training the machine learning models and for performing the real world performance evaluation are presented.

A. Machine Learning-enabled Data Rate Prediction

The training of the data rate prediction model is performed with the Waikato Environment for Knowledge Analysis (WEKA) [23] framework. Based on the findings and the open data sets of [4], the data rate prediction is performed with a RF regression model, which consists of 100 random trees and allows a maximum depth of 15. In order to consider MNO-specific characteristics, an individual prediction model is trained for each MNO and transmission direction.

Tab. I shows an overview of the coefficient of determination R^2, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) of the RF-based data rate prediction in uplink and downlink direction for the three considered MNOs.

Note that MAE and RMSE have to be considered with respect to the value range $S_{\max} - S_{\min}$ of the data set. As *MNO B* implements downlink carrier aggregation, it achieves a significantly higher value range – and absolute error measurements – than the other MNOs. A general observation is that the prediction works better in the uplink than in the downlink transmission direction. As the traffic intensity is typically much higher in the downlink than in the uplink [1],

Model	MNO A		MNO B		MNO C	
	UL	DL	UL	DL	UL	DL
R^2	0.779 ±0.023	0.588 ±0.021	0.678 ±0.04	0.634 ±0.062	0.718 ±0.028	0.493 ±0.036
MAE [MBit/s]	2.984 ±0.141	3.302 ±0.113	2.603 ±0.144	7.01 ±0.398	2.537 ±0.117	3.136 ±0.153
RMSE [MBit/s]	4.061 ±0.223	4.743 ±0.21	3.619 ±0.29	10.177 ±1.431	3.424 ±0.168	4.276 ±0.235
Range [MBit/s]	39.782	42.94	38.208	159.982	35.676	33.842

UL: Uplink, *DL*: Downlink
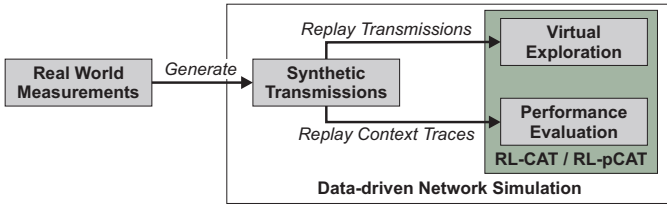
Fig. 4. Overview about the DDNS-based virtual exploration process and initial performance evaluation for parameter selection.

the resulting downlink data rate is highly impacted by the cell load, which can only be considered indirectly by means of the measurable RSRQ. In contrast to that, the uplink performance is more impacted by the network quality dynamics, which are represented by the whole network context feature set utilized by the prediction model.

B. Virtual Exploration Process

Although typical reinforcement learning techniques rely on a controlled trial-and-error mechanism, this method is unfavorable for the considered vehicular scenario as it would require to perform a multitude of real world drive tests in order to reach the convergence level of the proposed transmission schemes (see Sec. V-A). To overcome this issue, we apply a *virtual exploration process*, which is modeled within a DDNS setup (see Sec. II and [18]). An overview about the involved logical steps is provided in Fig. 4. Based on the available real world measurements of [4], we synthesized 2400 network context traces (time series data with two scenarios, periodic transmission interval $\Delta t = \{1, 2, ..., 120\ s\}$ with 10 repetitions) for three different MNOs in uplink and downlink direction. In total, the resulting data set consists of more than 84000 transmissions. The generated transmissions and context traces are replayed in random order with the proposed reinforcement learning-based transmission schemes according to Sec. III-A and Sec. III-B, which then learn to perform cognitive data transfer behaviors by identifying favorable and non-favorable transmissions from the previously performed actions. Moreover, the reinforcement learning approach extracts the complex interdependency between network quality and transmission results from the available data sets.

For the mobility-predictive RL-pCAT transmission scheme, the network context indicators – which are part of the feature vector of the data rate prediction model – are aggregated in a multidimensional connectivity map M which is jointly used with trajectory-aware mobility prediction according to [3]. However, a methodological dilemma needs to be solved as the closed loop scenario only allows to analyze the results of taken actions: If a transmissions is performed at time t, the transmission buffer is cleared and it cannot be directly concluded if the same transmission would have achieved a better performance at $t + \tau$. Therefore, the another DDNS evaluation is carried to analyze the behavior at $t + \tau$ and then update r'_{IDLE} at time t with Eq. 9.

C. Real World Performance Evaluation

After the DDNS-based exploration phase, the real world performance evaluation of the converged transmission schemes is carried out in the public cellular LTE networks of three different MNOs in Germany. Data is transmitted in uplink and downlink direction from the vehicle through the cellular network to a cloud-based server. A virtual sensor application generates 50 KB of sensor data per second which is buffered locally until the whole data buffer is transmitted and cleared. The resulting AoI of each successful transmission corresponds to the generation time of the oldest contained sensor packet. The measurement application is executed on Android-based User Equipments (UEs) (Samsung Galaxy S5 Neo, Model SM-G903F). We consider two different scenarios (suburban and highway) with different speed characteristics and building densities. For each of the tracks, 10 different drive tests are performed for each of the considered transmission schemes.

A summary about the parameters of the overall system is given in Tab. II. Further MNO-specific configurations are summarized in Tab. III.

Parameter	Value
Learning rate α	0.1
Context look ahead τ	10 s
Maximum buffering time $\Delta t_{\max}$	120 s
Trade-off factor w	**0.8**, 0, ..., 1.0
Deadline violation punishment Ω	-10
Connectivity map cell width c	25 m

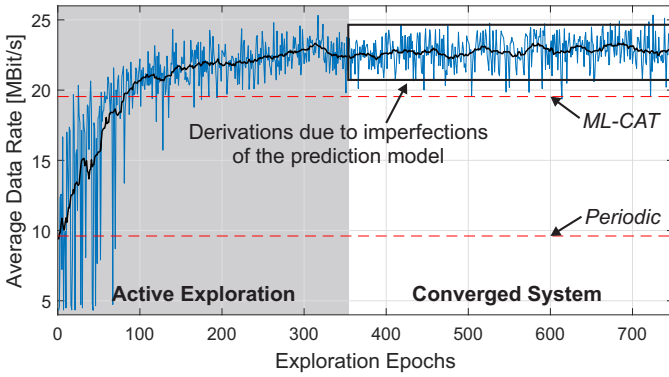| Model	MNO A		MNO B		MNO C	
	UL	DL	UL	DL	UL	DL
Target data rate $S*$	30	20	20	30	50	15
Maximum data rate $S_{\max}$	40	30	30	40	60	25

UL: Uplink, *DL*: Downlink

Fig. 5. Convergence of the virtual exploration process for uplink transmissions of MNO A. Each epoch corresponds to a single context trace of a DDNS evaluation. The black line shows the average behavior.
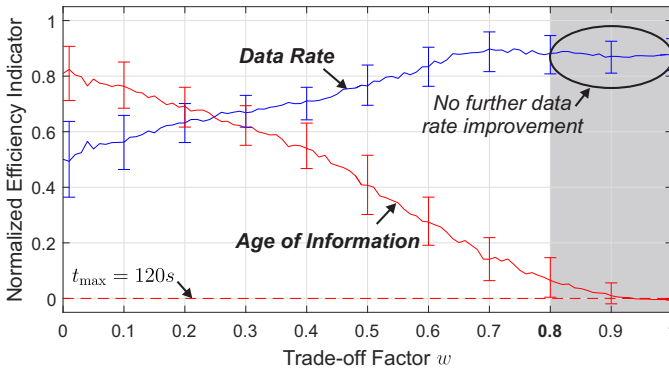

Fig. 7. Example temporal behavior of the RL-CAT transmission scheme. The flanks of the measurement trace show the actual transmission times.

Fig. 6. Trade-off between data rate efficiency $E_S = \bar{S}/S^*$ and age of information efficiency $E_{AoI} = 1 - \bar{\Delta}t/\Delta t_{max}$, which is controlled with the parameter w. The errorbar shows the standard deviation of the mean over 2400 evaluation runs per configuration.

(a) DDNS-enabled Reinforcement Learning

(b) RF-based Data Rate Prediction

Fig. 8. Temporal effort for training and applying the machine learning models.

V. RESULTS OF THE PERFORMANCE EVALUATION

In this section, the results for the simulative system optimization as well as for the real world performance evaluation are presented and discussed.

A. Exploration and System Optimization

At first, we investigate the required duration for the proposed methods to converge to a satisfying performance level.

The model convergence within the virtual exploration process is shown in Fig. 5 for the uplink data set of *MNO A*. Each epoch corresponds to a virtual single drive test synthesized by the DDNS, which contains the time series data of the context indicators as observed by the vehicle moving on its trajectory. For comparison, the baselines for periodic data transfer and the probabilistic ML-CAT approach are shown. After ∼80 epochs, the RL-CAT reliably outperforms the periodic approach and achieves a significant performance margin towards ML-CAT after around ∼200 epochs. Convergence is reached after ∼350 training epochs. It needs to be remarked that no perfect convergence can be achieved due to the remaining error range of the prediction model. Since the error-affected data rate prediction is the fundamental metric for the decision making,
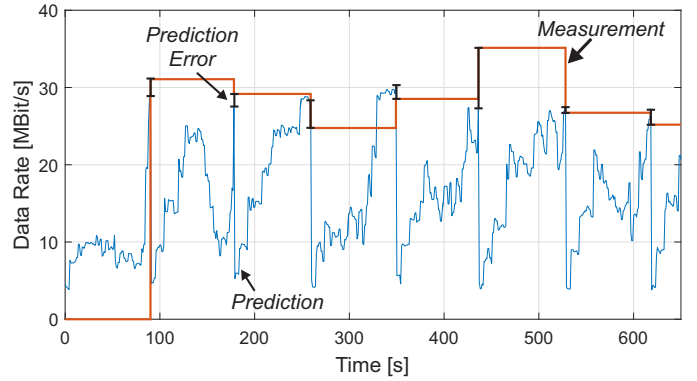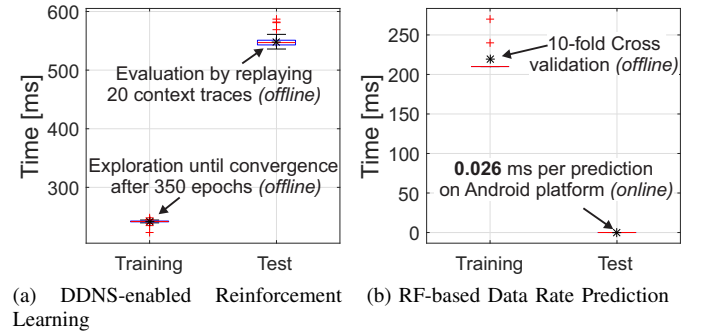
unfavorable decisions occur due to model imperfections. It can also be seen that the reinforcement learning process highly benefits from the computational efficiency of the virtual exploration with DDNS. Reaching convergence based on real world experiments would have required to perform more than 4000 km of drive tests. In contrast, DDNS does only require ∼250 ms to reach convergence by learning from replaying the transmissions of the 350 context traces (Intel Core i7-4770@3.4GHz platform with 32 GB RAM).

The parameter w allows to control the fundamental trade-off between data rate efficiency $E_S = \bar{S}/S^*$ and age of information efficiency $E_{AoI} = 1 - \bar{\Delta}t/\Delta t_{max}$ with $\bar{S}$ being the average data rate and $\bar{\Delta}t$ being the average AoI. For data rate optimization, the transmission scheme will rather prefer larger packet sizes in order to improve the payload-overhead ratio and to compensate the *slow start* mechanism of TCP. Fig. 6 shows the impact of the w for AoI and data rate. For $w > 0.8$, the age of information exceeds t_{max} for some transmissions, which results in a negative efficiency. In the following, we apply $w = 0.8$ in order to allow a performance comparison with CAT and ML-CAT, which focus on data rate optimization. Reduced transmission times lead to an early release of occupied resources and contribute to improving the intra-cell coexistence between multiple users [3].

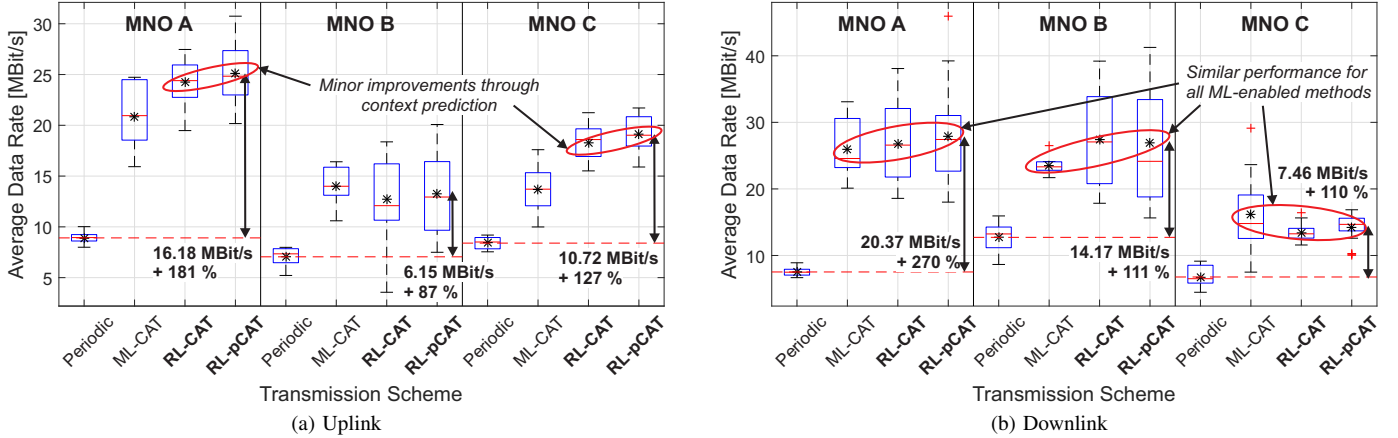Fig. 7 illustrates the temporal behavior of the RL-CAT

Fig. 9. Real world performance comparison of the resulting end-to-end data rates for the considered transmission schemes and MNOs.

transmission scheme with respect to the predicted data rate. The interdependency between payload size and achievable data rate can be clearly identified. After each transmission, the transmission buffer is cleared, which results in a drop of the predicted data rate due to a reduction of the payload-overhead-ratio. With in an increased buffering time and incoming sensor data, the predicted data rate is increased again.

The temporal effort related to the training and evaluation phases of the machine learning models is shown in Fig. 8.

Although the data-driven exploration process is performed based on a large amount of synthesized data, it only considers the actually performed transmissions. In contrast to that, the DDNS-based evaluation involves the replay of the whole time series data for each of the 20 context traces. Still, both parts can be processed rapidly. Online predictions on the Android platform have a practically negligible impact on the total execution time. Since the C++ implementation of the RF model consists of a binary tree of if/else conditions, it can be evaluated in real time.

B. Real World Performance Comparison

In the following, the converged transmission schemes are applied in the real world and compared to other transmission approaches. As references, we consider straightforward periodic transmission with a fixed interval $\Delta t = 10$ s and ML-CAT-based data transfer according to [3].

The overall results of the considered transmission schemes and MNOs are illustrated in Fig. 9 for uplink (a) and downlink (b) direction. As discussed in Sec. IV-A, the data rate prediction works more accurately in the uplink as the cell is more impacted by channel-related effects than by congestion. A general observation is the the that the predictive method RL-pCAT achieves slight improvement compared to RL-CAT, which outperforms the other approaches in most cases. In comparison to periodic data transfer, a data rate improvement by 181% is achieved in the uplink and by 270% in the downlink for *MNO A*. For *MNO B*, the uplink data rate prediction is not very accurate ($R^2 = 0.678 \pm 0.04$), which leads to a slightly worse performance for RL-CAT than for ML-CAT. However, RL-pCAT is able to compensate many of the outliers through its context-predictive behavior. In the downlink transmission direction, all opportunistic approaches achieve a similar level of improvement compared to periodic transfer. For *MNO A* and *MNO B*, the proposed reinforcement learning-based approaches outperform the other opportunistic methods. *MNO C* suffers from a low downlink data rate prediction accuracy.

Since the reinforcement-based decision making is based on the predicted data rate, future optimizations of the proposed
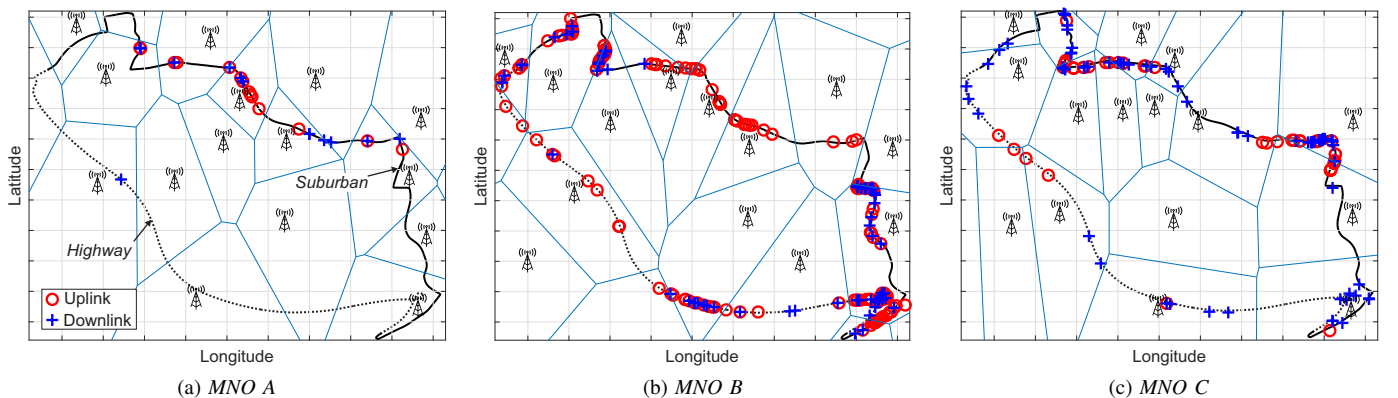

Fig. 10. Geospatial distribution of the blackspot regions where the transmissions significantly deceed the average data rate of the MNO ($S(t) < \bar{S}/2$). The voronoi diagram shows the MNO-specific eNBs locations.

scheme should aim to increase the prediction accuracy. A promising approach is the application of *cooperative* approaches for cell load estimation. Upcoming 5G networks explicitly consider machine learning-based load analysis through a Network Data Analytics Functions (NWDAF) [24]. Although this method is a part of the core network, providing the acquired information for the UEs – e.g., via the control channels – could lead to significant improvements for client-side context-aware data transfer.

Many of the reasons for significant prediction errors are related to geospatial effects such as *cellular handovers* and even technology fallbacks. In road safety management, the term *blackspot* refers to regions with a high probability for road accidents. Analogously, we can define *communication blackspots* which show a clustering of low data rate transmissions.

Fig. 10 provides a map of the two evaluation scenarios and the network infrastructure locations of all considered MNOs. Furthermore, all transmission that fulfill $S(t) < \bar{S}/2$ and therefore significantly deceed the average transmission performance of the MNO are shown. It can be seen that those transmission can be aggregated to blackspot regions for each of the MNO. Although the voronoi diagram only considers the eNB locations and not the real resulting coverage areas, many of the blackspot regions are close to the cell borders. It is very plausible that future opportunistic methods can achieve further improvements by proactively considering blackspot regions within the reinforcement learning-based transmission process.

VI. Conclusion

In this paper, we presented a reinforcement learning-based transmission approach for optimizing the end-to-end performance of vehicular data transfer. The proposed opportunistic communication scheme schedules transmissions cognitively with respect to the predicted channel conditions.

The results of the real world performance evaluation show that the proposed approach significantly outperforms existing probabilistic channel-aware transmission schemes in most scenarios and is able to achieve massive improvements in the resulting data rate compared to typically considered periodic data transfer.

In future work, we will optimize the data rate prediction accuracy by explicitly considering blackspot context information and by applying a network-assisted load estimation approach similar to NWDAF. On this foundation, we will furthermore develop a reinforcement learning based transmission scheme for multi-MNO networks and investigate the performance of *multi-armed bandits* and *Deep Reinforcement Learning (DRL)*.

References

[1] 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 & Tutorials*, 2017.

[2] B. Sliwa, T. Liebig, R. Falkenberg, J. Pillmann, and C. Wietfeld, "Efficient machine-type communication using multi-metric context-awareness for cars used as mobile sensors in upcoming 5G networks," in *2018 IEEE 87th Vehicular Technology Conference (VTC-Spring)*, Porto, Portugal, Jun 2018, Best Student Paper Award.

[3] B. Sliwa, R. Falkenberg, T. Liebig, N. Piatkowski, and C. Wietfeld, "Boosting vehicle-to-cloud communication by machine learning-enabled context prediction," *IEEE Transactions on Intelligent Transportation Systems*, Jul 2019.

[4] B. Sliwa and C. Wietfeld, "Empirical analysis of client-based network quality prediction in vehicular multi-MNO networks," in *2019 IEEE 90th Vehicular Technology Conference (VTC-Fall)*, Honolulu, Hawaii, USA, Sep 2019.

[5] B. Sliwa, "Raw data of real world measurements," Oct 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3490335

[6] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, April 2017.

[7] H. Ye, L. Liang, G. Y. Li, J. Kim, L. Lu, and M. Wu, "Machine learning for vehicular networks: Recent advances and application examples," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 94–101, June 2018.

[8] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, 2nd ed. The MIT Press, 2018.

[9] H. Gacanin, "Autonomous wireless systems with artificial intelligence: A knowledge management perspective," *IEEE Vehicular Technology Magazine*, pp. 1–1, 2019.

[10] T. Mangla, N. Theera-Ampornpunt, M. Ammar, E. Zegura, and S. Bagchi, "Video through a crystal ball: Effect of bandwidth prediction quality on adaptive streaming in mobile environments," in *Proceedings of the 8th International Workshop on Mobile Video*, ser. MoVid '16. New York, NY, USA: ACM, 2016, pp. 1:1–1:6.

[11] F. Bouali, K. Moessner, and M. Fitch, "A context-aware user-driven framework for network selection in 5G multi-RAT environments," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sep. 2016, pp. 1–7.

[12] M. Akselrod, N. Becker, M. Fidler, and R. Luebben, "4G LTE on the road - what impacts download speeds most?" in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–6.

[13] J. Riihijarvi and P. Mahonen, "Machine learning for performance prediction in mobile cellular networks," *IEEE Computational Intelligence Magazine*, vol. 13, no. 1, pp. 51–60, Feb 2018.

[14] F. Jomrich, A. Herzberger, T. Meuser, B. Richerzhagen, R. Steinmetz, and C. Wille, "Cellular bandwidth prediction for highly automated driving - Evaluation of machine learning approaches based on real-world data," in *Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems 2018*, no. 4. SCITEPRESS, Mar 2018, pp. 121–131.

[15] A. Samba, Y. Busnel, A. Blanc, P. Dooze, and G. Simon, "Instantaneous throughput prediction in cellular networks: Which information is needed?" in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, May 2017, pp. 624–627.

[16] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 5 2015.

[18] B. Sliwa and C. Wietfeld, "Towards data-driven simulation of end-to-end network performance indicators," in *2019 IEEE 90th Vehicular Technology Conference (VTC-Fall)*, Honolulu, Hawaii, USA, Sep 2019.

[19] E. R. Cavalcanti, J. A. R. de Souza, M. A. Spohn, R. C. d. M. Gomes, and A. F. B. F. d. Costa, "VANETs' research over the past decade: Overview, credibility, and trends," *SIGCOMM Comput. Commun. Rev.*, vol. 48, no. 2, pp. 31–39, May 2018.

[20] J. Lee, J. Lee, Y. Im, S. Dhawaskar Sathyanarayana, P. Rahimzadeh, X. Zhang, M. Hollingsworth, C. Joe-Wong, D. Grunwald, and S. Ha, "CASTLE over the air: Distributed scheduling for cellular data transmissions," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '19. New York, NY, USA: ACM, 2019, pp. 417–429.

[21] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, May 1992.

[22] B. Sliwa, T. Liebig, R. Falkenberg, J. Pillmann, and C. Wietfeld, "Machine learning based context-predictive car-to-cloud communication using multi-layer connectivity maps for upcoming 5G networks," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Chicago, USA, Aug 2018.

[23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[24] 3GPP, "5G System; Network Data Analytics Services;Stage 3," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 29.520, Mar 2019, version 15.3.0.