

Diversity Maximized Scheduling in RoadSide Units for Traffic Monitoring Applications

Ahmad Sarlak*, Abolfazl Razi*, Xiwen Chen*, and Rahul Amin†

*School of Computing, Clemson University, Clemson, SC, USA

†Tactical Networks Group, MIT Lincoln Laboratory, Lexington, MA, USA

Emails: {asarlak, arazi,xiwenc}@clemson.edu, rahul.amin@ll.mit.edu

Abstract—This paper develops an optimal data aggregation policy for learning-based traffic control systems based on imagery collected from Road Side Units (RSUs) under imperfect communications. Our focus is optimizing semantic information flow from RSUs to a nearby edge server or cloud-based processing units by maximizing data diversity based on the target machine learning application while taking into account heterogeneous channel conditions and constrained total transmission rate. To this end, we enforce fairness among class labels to increase data diversity for classification problems. Furthermore, we propose a greedy interval-by-interval scheduling policy powered by coalition game theory to reduce the computation complexity. Once, RSUs are selected, we employ a maximum uncertainty method to handpick data samples that contribute the most to the learning performance. Our method yields higher learning accuracy compared to random selection, uniform selection, and network-based optimization methods (e.g., FedCS)¹.

I. INTRODUCTION

Road-Side Units (RSUs) are an integral part of smart transportation systems due to their role in collecting visual information and broadcasting traffic-advisory commands [1]. With the rise of Edge Computing (EC), the bulk of heavy computations can be performed in edge servers located in RSUs or the entry point of the wireless network in the vicinity of RSUs [1]. RSUs can collect data for developing universal models in centralized servers [2]–[4]. It is known that the limitations of wireless networks can influence the quality of the learning-based analysis and decision-making platforms [5], [6]. A large body of work is devoted to characterizing the impacts of networking factors on the quality of Distributed Deep Learning (DDL) tasks as well as enhancing networking performance to improve the ultimate quality of DDL [7]–[9].

Another closely-related line of research is adapting data-sharing and model-sharing strategies based on the limitations of the underlying network. We consider this issue from a substantially different perspective by regulating packet transmission under imperfect networking so that the *diversity* of accumulated samples in the processing unit is maximized, noting that high diversity typically translates to higher performance. More specifically, we aim to optimize data collection from a set of RSUs with heterogeneous networking conditions when the total data aggregation limit is constrained. Our goal is to optimize *semantic information exchange*, which may not necessarily be equivalent to optimizing the raw data throughput. This approach is driven by the fact that some

data samples may not significantly contribute to the ultimate learning quality [10].

To this end, we optimize the semantic diversity of the collected data by enforcing *fairness* among data classes to increase data diversity. Our approach is different than pure communication-based optimization methods, where the goal is optimizing communication performance metrics such as delay and throughput without considering the semantic content of data packets. Simultaneously, our perspective is different than typical fairness-imposing scheduling methods that try to balance resource utilization by different network nodes without considering their contribution to the performance of the target ML application [11]. Our contribution can be summarized as i) optimized scheduling for imperfect communications while maintaining high diversity by imposing fairness among accumulated class labels, ii) using coalition game theory to characterize the added data diversity by any selection of RSUs in an interval-by-interval fashion, and iii) using constrained satisfaction problem to translate optimal attempt probabilities to a binary scheduling matrix.

II. SYSTEM MODEL

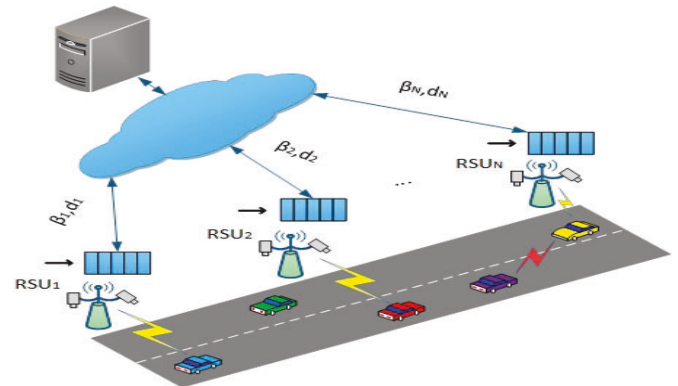


Fig. 1: Roadside Units collect data samples from vehicles and send them to a central cloud-based computation server for learning-based processing.

We assume that there are N RSUs represented by $n_1, n_2, \dots, n_N$ in a specific traffic zone. Each RSU, equipped with an Edge Computing (EC) server, pre-processes the collected imagery, then exploits training samples and sends them to a central processing unit through wireless links. There exist $M < N$ channels with an equal bandwidth $b_1 = b_2 = \dots b_M = bw$, so the number of simultaneous packets by all N RSUs can not exceed M . To be more specific, suppose that a transmission cycle (we also call it interval)

¹This material is based upon the work supported by the National Science Foundation under Grant Numbers 2008784 and 2204721.

includes T timeslots. We present the transmission matrix as a $Q_{N \times T} = [Q_{it}]$, where $Q_{it} = 1$ means that RSU n_i sends a packet at timeslot t and is silent otherwise.

The goal of the scheduling is to fill in the binary scheduling matrix (interval by interval) so that the number of simultaneous transmissions in each timeslot does not exceed M , meaning that $\sum_{i=1}^N Q_{it} \leq M, \forall t \in T$. We can solve this problem in two sequential steps. First, we define $\alpha_i = \frac{1}{T} \sum_{t=1}^T Q_{it}$ for RSU n_i , which represents the transmission attempt probability in a stationary case (during one transmission cycle: T timeslots). Then, the optimization would reduce to finding the attempt probability vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ so that

$$\begin{aligned} \sum_{i=1}^N \alpha_i &= \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T Q_{it} \right) \\ &= \frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^N Q_{it} \right) \leq \frac{1}{T} \sum_{t=1}^T M = M, \quad \forall i \in N. \end{aligned} \quad (1)$$

Once, we obtain attempt probabilities α_i (the row weights of Q), then the next step would be to permute $\alpha_i T$ ones and $(1 - \alpha_i)T$ zeros in row i so that $\sum_{i=1}^N Q_{it} \leq M$ is satisfied. We use constrained satisfaction problem by considering rows $r_i = (Q_{i1}, Q_{i2}, \dots, Q_{iT})$ as variables and all permutations of $\lfloor \alpha_i T \rfloor$ ones and $\lceil (1 - \alpha_i)T \rceil$ zeros as values, where we used $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ to round up/down to an integer value, and select T to be the minimum product of $\lfloor \alpha_i M \rfloor$ terms. We will perform iterative assignment by random inconsistent variable selection and min-conflict heuristic for value selection until constraint $\sum_{i=1}^N Q_{it} \leq M$ is satisfied. This will fully determine transmission matrix Q with row weights $\alpha_i T$ and column vectors bounded by M .

III. PERFORMANCE METRICS

The problem is formulated as follows

$$\begin{aligned} \arg \max_{\alpha} f(\alpha, t) &= \sum_{i=1}^3 w_i f_i(\alpha_1, \alpha_2, \dots, \alpha_N, t) \\ \text{s.t.} \quad &0 \leq \alpha_i \leq 1, \quad \text{and} \quad \sum_{i=1}^N \alpha_i \leq M, \end{aligned} \quad (2)$$

where the objective function is the linear combination of the set of desired performance metrics. In our case, $f_1(\alpha, t)$, $f_2(\alpha, t)$, and $f_3(\alpha, t)$, respectively, represent the average delay, throughput, and diversity metrics at time (interval) t . Here, the throughput and delay terms $f_1()$ and $f_2()$ are calculated per interval but the diversity term $f_3()$ depends on the RSU's transmitted packets in the previous timeslot, which hinders solving the problem in an interval-by-interval fashion. The optimization is constrained by $0 \leq \alpha_i \leq 1$ meaning that each RSU can not utilize more than one channel at a time and $\sum_{i=1}^N \alpha_i \leq M$ meaning that the central unit can not receive more than M packets simultaneously. To have a non-trivial solution, we set $M < N$.

A) Delay: We take a simplistic assumption and model $D_i^{(1)}$ the E2E delay for one packet for RSU n_i as an exponentially distributed continuous-valued Random Variable (RV) with RSU-specific mean λ_i , and then incorporate the impact of re-transmissions. Specifically, we define $D_i^{(1)} \sim f_{\lambda_i}(d_i)$

$$f_{\lambda_i}(d_i) = \begin{cases} \lambda_i e^{-\lambda_i d_i}, & d_i \geq 0, \\ 0, & d_i < 0. \end{cases}$$

We consider that an average delay $E[D_i^{(1)}] = 1/\lambda_i$ for each RSU n_i remains constant during one scheduling interval (one optimization round). λ_i can be drawn from Gamma distribution [12] in an interval-by-interval fashion. Here, we consider $D_i^{(1)}$ captures all delay terms. However, if it accounts only for queuing delay, we can add the constant term $D_{tr} = L/R_{ch} = L/\alpha b_w$ accounting for the actual transmission delay, where L is the packet length (in bits), b_w is the bandwidth (in Hz) and $R_{ch} = \alpha b_w$ is the transmission rate (in bits/sec) for a given transmission system.

This delay $D_i^{(1)}$ is for one packet transmission; hence, the effective E2E delay D_i should be calculated for the last attempt when re-transmissions are involved. During one transmission interval, the number of re-transmissions R_i for RSU n_i follows a geometric distribution. Specifically, we have $P(R_i = r) = (1 - \beta_i)^{r-1} \cdot \beta_i$ where β_i is the packet drop probability for RSU n_i drawn interval-by-interval from a Beta distribution with shape and scale parameters $\beta_i \sim \text{Beta}(a, b)$. Therefore, the expected value of E2E delay for one successful transmission is calculated as follows

$$\begin{aligned} E[D_i] &= E_R[E_D[D_i | R]] \\ &= E_R[E[\underbrace{D_i^{(1)} + D_i^{(1)} + \dots + D_i^{(1)}}_{R \text{ times}}]] \\ &= E_R[R \cdot E[D_i^{(1)}]] = E[R] \cdot E[D_i^{(1)}] = (1/\lambda_i) \cdot 1/(1 - \beta_i), \end{aligned} \quad (3)$$

where we used $E[R] = 1/(1 - \beta_i)$ and $E[D_i^{(1)}] = 1/\lambda_i$. The average delay of the entire system is apparently the linear combination of RSU-specific delays weighted by their attempt probabilities. We consider the negative/inverse of the average delay as our first term in the objective function in (2). Therefore, we have

$$f_1(\alpha, t) = 1 / \sum_{i=1}^N \alpha_i(t) E[D_i(t)] = 1 / \sum_{i=1}^N \frac{\alpha_i(t)}{(\lambda_i(t))(1 - \beta_i(t))} \quad (4)$$

at transmission interval t . Note that we drop t from some equations when it is clear from the context.

B) Throughput: Likewise, we can calculate the effective throughput ζ_i of RSU n_i as $\zeta_i = \alpha_i R_{ch} / E[R_i] = \alpha_i R_{ch} (1 - \beta_i)$ where R_{ch} is the rate of channel, $E[R_i] = 1/(1 - \beta_i)$ is the average transmission per packet. The second term in the objective function of (2) is the system throughput (at time t), which is simply the sum of the individual throughputs. We use

$$f_2(\alpha, t) = \sum_{i=1}^N \zeta_i(t) = \sum_{i=1}^N \alpha_i(t) R_{ch} (1 - \beta_i(t)). \quad (5)$$

C) Diversity through Fairness: Collected data from RSUs can vary greatly due to factors like illumination, camera settings, road conditions, and traffic composition. It is known that the diversity of data samples significantly enhances learning quality [13], [14], [14] and may conflict with the efficiency

of data accumulation. we enforce diversity through fairness among collected data samples of different classes which we can view it as *soft fairness*. In this work, we use Jain's fairness index, defined as $J(\mathbf{x}) = \frac{(\sum_{i=1}^N x_i)^2}{N \sum_{i=1}^N x_i^2}$. To enhance diversity, we impose fairness among data categories (e.g., class labels in multi-level classification). Suppose c_i^j is the number of data samples of class j in RSU n_i . Then, the diversity of RSU n_i is represented by $\mathbf{c}_i = [c_i^1, c_i^2, \dots, c_i^C]$. Assuming that samples are selected at random by RSUs, then the number of received samples of class j until time t , $c_R^j(t)$ is the sum of samples of the same class attempted by all RSUs in all transmission intervals up to time t proportional to their effective throughput as $c_R^j(t) = \sum_{\tau=0}^t \sum_{i=1}^N c_i^j \zeta_i(\tau)$ where $\zeta_i(\tau)$ is the throughput of RSU n_i at time τ given by ζ_i . We impose fairness on the received data samples of all classes as

$$f_3(\boldsymbol{\alpha}, t) = J(c_R^1(t), c_R^2(t), \dots, c_R^C(t)). \quad (6)$$

IV. COALITION-BASED GREEDY SCHEDULING

The optimization problem in (2) is non-convex, hence does not admit a closed-form solution or KKT approach. If we divide the range of alpha $[0 \ 1]$ into $[0, d\alpha, 2d\alpha, \dots, 1]$ with N_α steps $d\alpha = 1/N_\alpha$, the computation would be in the orders of $(N_\alpha)^N$ considering only $O(1)$ complexity to evaluate objective functions $f_i(\boldsymbol{\alpha})$. Therefore, it is NP-hard in the number of RSUs and can be prohibitively expensive for large-scale systems. Moreover, it does not allow interval-by-interval optimization, because the fairness index $f_3()$ should account for all preceding transmission intervals.

To address this issue, we propose an approximate method to select top- K RSUs using *coalition game theory*. Game theory is an appropriate tool to evaluate fairness since it quantifies the contribution of each player n when joining a coalition S , as *marginal value* $v(S \cup \{n\}) - v(S)$, where $v(S)$ is the value function representing the total payoff can be gained by the members of coalition S . Here we define

$$\begin{aligned} v(S(t)) &= \sum_{i=1}^3 w_i f_i(\alpha_1, \alpha_2, \dots, \alpha_N, t) \\ &= w_1 / \sum_{n \in S(t)} \frac{\alpha_n(t)}{(\lambda_n(t))(1 - \beta_n(t))} \\ &\quad + w_2 \sum_{n \in S(t)} \alpha_n(t) R_{ch}(1 - \beta_n(t)) \\ &\quad + w_3 J(c_R^1(t), c_R^2(t), \dots, c_R^C(t)) \end{aligned} \quad (7)$$

$$\alpha_n = \begin{cases} M/|S(t)| & n \in S(t) \\ 0 & \text{else} \end{cases} \quad (8)$$

representing a case where coalition members split M transmission resources equally. In standard coalition games, we can have coalitions of arbitrary size. Also, Shapley value of each player is defined as the expected marginal contribution of player i to the set of players who precede this player as

$$\phi_i(v) = \frac{1}{N!} \sum_{S \subseteq N/i} \frac{|S|!(N - |S| - 1)!}{N!} [v(S \cup \{n_i\}) - v(S)] \quad (9)$$

This involves evaluating the value of all 2^N coalitions. Here, to reduce complexity, we allow only the formation of fixed-size coalitions $|S(t)| = K$ as the set of active RSUs in transmission interval t . Apparently, we must have $|S| \geq M$ to maintain constraint $\alpha_i = M/|S| < 1$.

V. UNCERTAINTY-BASED SAMPLE SELECTION

Once we determine active RSUs by solving (2) and (7), we can select a balanced number of samples among different classes. However, there is flexibility in selecting samples within each class. Our approach to this problem is sensing samples that can contribute the most to the ML application at hand. For instance, for multi-level classification, we use the *min-margin* criteria by selecting samples with the lowest difference between the softmax highest and second highest probability [15]. To this end, the Fusion Center updates the model at the end of each transmission interval and sends back the model parameters. The RSU selects the samples that exhibit maximum uncertainty.

VI. SIMULATION

In this section, we investigate the performance of the proposed interval-by-interval scheduling policy in terms of the ultimate learning quality under time-varying conditions. We use the following two datasets for our experiments. The first dataset is [16], in which we apply traffic light status detection (red, yellow, green) with a convolutional neural network (CNN). We investigate scenarios where data samples are split unequally (in terms of class labels) among RSUs to represent *unbalanced* datasets. We evaluated the following methods for each test scenario: i) optimized scheduling without fairness by setting $w_3 = 0$ in Eqs (2) and (7) to exclude fairness, ii) optimized scheduling with fairness, where we set $w_1 = w_2 = w_3$ to weight delay, throughput, and fairness equally (after applying proper normalization for each metric to be in the same scale with zero-mean and unit variance), iii) uniform rate, where all RSUs utilize an equal number of resources, and we set $\alpha_1 = \alpha_2 = \dots = \alpha_N = M/N$, iv) random rate, where $K = 5$ out of $N = 10$ RSUs are selected in random to transmit their packets with attempt probability $\alpha_i = M/K$, v) FedCS as an exemplary communication-based method [17].

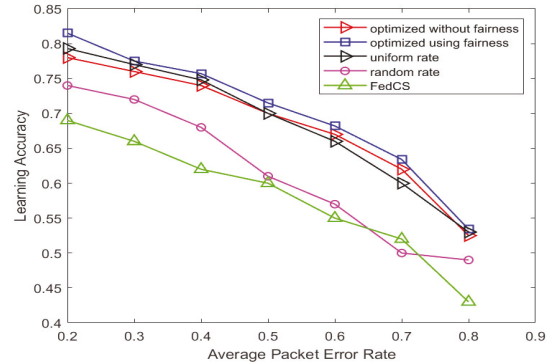


Fig. 2: Learning accuracy of the system using different scheduling policies versus packet error rate. The results are for unbalanced CIFAR-10 dataset among RSUs. We set $N = 10$, $M = K = 5$, and $T = 100$.

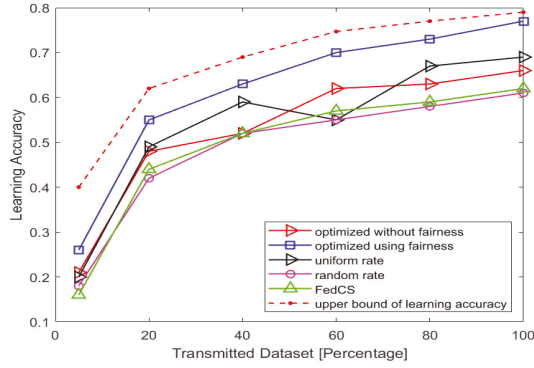


Fig. 3: Online learning accuracy of the system using different scheduling policies for different portions of transmitted samples. The results are for an unbalanced CIFAR-10 dataset among RSUs. We set $N = 10$, $M = K = 5$, and $T = 100$.

We evaluate our method by training a CNN on CIFAR10 dataset data samples sent by RSUs under different scheduling policies. We consider unbalanced datasets, where ten classes split inequality among RSUs. Indeed, to simulate an unbalanced dataset, we include only 2 out of 10 classes in each RSU dataset. The results are shown in Figs. 2 and 3. It is seen in Fig. 2 that the ultimate learning accuracy gradually increases by sending more packets consistently for the optimized scheduling by enforcing fairness on unbalanced datasets. Fig. 2 demonstrates a decline in the accuracy by increasing the PDR. In this figure, the proposed optimization when enforcing fairness outperforms all methods for an unbalanced dataset (Figs. 2 and 3), since it increases the throughput and the optimized scheduling with enforcing diversity yields superior performance. This highlights the fact that optimizing the scheduling policy merely based on the networking parameters is not optimal for learning-based applications.

TABLE I: Evaluation with traffic dataset. Three RSUs are used with $\beta_1 = 0.1$, $\lambda_1 = 1.3$, $\beta_2 = 0.22$, $\lambda_2 = 1.5$, $\beta_3 = 0.44$, $\lambda_3 = 1.1$.

Methods	Channel Utilization	Throughput	Learning Quality
Optimize without fairness	RSU1 = 0.58 RSU2 = 0.3 RSU3 = 0.12	89	0.85
Optimize with fairness	RSU1 = 0.42 RSU2 = 0.32 RSU3 = 0.21	70	0.93
Uniform rate	RSU1 = 0.33 RSU2 = 0.33 RSU3 = 0.33	60	0.82
Random rate	RSU1 = 0.28 RSU2 = 0.44 RSU3 = 0.17	46	0.81
FedCS	RSU1 = 0.45 RSU2 = 0.5 RSU3 = 0.07	59	0.78

Similar results are provided for the traffic light dataset in Table I. It can be observed that FedCS favors RSU2 for its lower expected delay $1/\lambda_2 = 1/1.5$. On the other hand, the optimized method favors RSU1 for its lower packet drop rate $\beta_1 = 0.1$. Nevertheless, the attempt probabilities are slightly different with and without enforcing diversity. The optimized scheduling with no diversity maximizes the system throughput whereas the optimized policy with enforcing diversity through fairness among the class labels of the received samples results in the highest learning accuracy of 87%.

VII. CONCLUSION

We investigated the importance of enforcing diversity among collected data samples from RSUs for traffic monitoring applications. We offered a new coalition-based greedy optimization that enforces the diversity of the received dataset by imposing fairness on the collected class labels in an interval-by-interval fashion. Then, we used the min-margin criterion to select samples from each class that are less consistent with the trained learning system (hence contributing the most to improving it). This improves upon the current practice of optimized scheduling merely based on networking parameters. Particularly, our method outperforms random scheduling, uniform scheduling, and communication-based scheduling methods by a significant margin in terms of learning quality (more than 5% improvement in classification rate).

REFERENCES

- [1] X. Chen, H. Wang, A. Razi, B. Russo, J. Pacheco, J. Roberts, J. Wishart, L. Head, and A. G. Baca, "Network-level safety metrics for overall traffic safety assessment: A case study," *IEEE Access*, 2022.
- [2] A. Mchergui, T. Moulahi, and S. Zeadally, "Survey on artificial intelligence (ai) techniques for vehicular ad-hoc networks (vanets)," *Vehicular Communications*, vol. 34, p. 100403, 2022.
- [3] A. Razi, X. Chen, H. Li, H. Wang, B. Russo, Y. Chen, and H. Yu, "Deep learning serves traffic safety analysis: A forward-looking review," *IET Intelligent Transport Systems*, 2022.
- [4] S. P. H. Boroujeni and E. Pashaei, "Data clustering using chimp optimization algorithm," in *2021 11th international conference on computer engineering and knowledge (ICCKE)*, pp. 296–301, IEEE, 2021.
- [5] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269–283, 2020.
- [6] M. Chinipardaz, S. M. Noorhosseini, and A. Sarlak, "Inter-cell interference in multi-tier heterogeneous cellular networks: modeling and constraints," *Telecommunication Systems*, vol. 81, no. 1, pp. 67–81, 2022.
- [7] J. Konečný, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," *arXiv preprint arXiv:1511.03575*, 2015.
- [8] F. Lotfi, O. Semiari, and W. Saad, "Semantic-aware collaborative deep reinforcement learning over wireless cellular networks," in *ICC 2022-IEEE International Conference on Communications*, pp. 5256–5261, IEEE, 2022.
- [9] A. Sarlak and Y. Darmani, "An approach to improve the quality of service in dtn and non-dtn based vanet," *Journal of Information Systems and Telecommunication (JIST)*, vol. 4, no. 32, p. 240, 2021.
- [10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.
- [11] C. Chaieb, F. Abdelkefi, and W. Ajib, "Deep reinforcement learning for resource allocation in multi-band and hybrid ota-noma wireless networks," *IEEE Transactions on Communications*, 2022.
- [12] N. A. Ali, E. Ekram, A. Eljasmy, and K. Shuaib, "Measured delay distribution in a wireless mesh network test-bed," in *2008 IEEE/ACS International Conference on Computer Systems and Applications*, pp. 236–240, IEEE, 2008.
- [13] A. Kulesza, B. Taskar, et al., "Determinantal point processes for machine learning," *Foundations and Trends® in Machine Learning*, vol. 5, no. 2–3, pp. 123–286, 2012.
- [14] M. Soleimani, F. Mahmudi, and M. Naderi, "Some results on the maximal graph of commutative rings," *Advanced Studies: Euro-Tbilisi Mathematical Journal*, vol. 16, no. 1, pp. 21–26, 2023.
- [15] T. Scheffer, C. Decomain, and S. Wrobel, "Active hidden markov models for information extraction," in *International Symposium on Intelligent Data Analysis*, pp. 309–318, Springer, 2001.
- [16] M. B. Jensen, M. P. Philipsen, A. Møgelmoose, T. B. Moeslund, and M. M. Trivedi, "Vision for looking at traffic lights: Issues, survey, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 1800–1815, 2016.
- [17] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *ICC 2019-2019 IEEE international conference on communications (ICC)*, pp. 1–7, IEEE, 2019.