Fog-Based Data Offloading in UWSNs with Discounted Rewards: A Contextual Bandit

Yuchen Shan, Hui Wang(✉), Zihao Cao, Yujie Sun, and Ting Li

School of Mathematics and Computer Science, Zhejiang Normal University, Jinhua, People's Republic of China
hwang@zjnu.cn

Abstract. Urban wireless sensor networks (UWSNs) are an important application scenario for the Internet of Things (IoT). Nevertheless, applications based on urban environments are often computationally intensive, and sensor nodes are resource-constrained and heterogeneous. Fog computing has the potential to liberate the computation-intensive mobile nodes through data offloading. Therefore, reliable data collection and scalable coordination based on fog computing are seen as a challenge. In this paper, the challenge of data offloading is modeled as a contextual bandit problem—an important extension of the multi-armed bandit. First, the heterogeneity of the sensor nodes is used as contextual information, allowing the network to complete data collection at a small computational cost. Second, an ever-changing environmental scenario is considered in which the distribution of re-wards is not fixed, but varies over time. Based on this non-stationary bandit model, we propose a contextual bandit algorithm NCB-rDO in order to improve the success rate of data offloading, which solves the problem of data loss when the contextual information changes suddenly. Experimental results demonstrate the effectiveness and robustness of this data offloading algorithm.

Keywords: Contextual bandit · Collaborative offloading · Dynamic fog computing · Urban wireless sensor network

1 Introduction

Our lives are increasingly dependent on smart IoT devices that collect data from the environment and operate in the physical world [1]. Many sensor nodes are deployed in cities and large urban metropolitan areas. Examples include security monitoring, traffic, pollution monitoring, infotainment, energy management [2, 3]. The proliferation of urban IoT applications has created an unprecedented amount of data, including physical quantities sampled from the environment [4]. Due to the limited resources of the sensor, it is generally not possible to process sampled data locally and often relies on the cloud for data analysis and long-term storage [5]. However, this will inevitably result in additional energy consumption and computational overhead for offloading the data. Fog network distributes computing, storage, control, and communication services across the cloud-to-thing continuum [6], rather than offloading them to the cloud. As a result, data

H. Gao et al. (Eds.): ChinaCom 2021, LNICST 433, pp. 509–522, 2022.
https://doi.org/10.1007/978-3-030-99200-2_38

collected by IoT sensors from the environment is collected by mobile fog nodes, such as cars or smartphones. In attempting to meet the performance requirements of particular applications [7], the fog network considers the resources available on the devices, such as transmission rate, efficiency, and network cost. Thus, the fog-based approach could exploit the dynamic heterogeneity of the sensors, which is caused by the ever-changing urban environment.

Sensor nodes can only effectively support urban applications if they first successfully transmit data to the fog nodes. Nevertheless, the limited storage resources of these sensors prevent them from keeping the data they generate indefinitely. It is therefore vital that data is transferred to resource-rich nodes before storage space is exhausted. We refer to this problem as the "data loss problem". But, designing solutions to prevent data loss is very challenging: an effective solution requires careful cooperation between sensor nodes and fog nodes, but the limited computation of sensors precludes coordination mechanisms with significant overhead. Moreover, the dynamic heterogeneity of nodes such as their memory margin, transmission rate, etc. cannot be controlled or predicted. Finally, in the case of sudden changes in node heterogeneity, nodes with as high a quality as possible must be selected for offloading data.

Recently, the multi-armed bandit (MAB) has been used to solve wireless communication and network decision problems [8, 9]. In the stochastic MAB problem, given a set of arms (actions), one arm is selected on each trial and a reward is obtained from the reward distribution followed by that arm. Each arm has an unknown random reward, and by pulling an arm, the player gets an immediate reward. Players decide which arm to pull in a series of trials to maximize the rewards that accrue over time. To extend the MAB framework to dynamic complex systems with contextual relevance, the contextual bandit (CB) model, which is widely used in recommender systems [10], was investigated. Unlike MAB, the rewards in the CB model depend on the contextual information provided at each period. We attempt to design the low-complexity data offloading problem as a CB problem and use the dynamic heterogeneity of nodes, i.e., state information, as contextual information. Furthermore, given the dynamics of the urban environment, the data offloading problem can be constructed as a non-stationary MAB problem. In this non-stationary case, the reward distribution can change at any moment in time at an unknown moment. Thus, a discount factor was introduced to treat historical rewards differently. To estimate instantaneous expected rewards, the data offloading scheme with discounted rewards averaged past rewards and gave a discount factor that gave greater weight to the most recent observations.

The motivation of this paper is a fog-based architecture that takes into account the sudden change in dynamic heterogeneity of sensor nodes due to the external environment (cold, earthquakes, etc.) and selects which nodes have a higher quality (good state) that can improve the success of data offloading and thus avoid the data loss problem. In this paper, sensors in the environment are logically divided into two categories: the sensors that offload data are called task nodes, while the sensors that receive data are called helper nodes. The kernel UCB algorithm [11] is a non-linear algorithm based on the CB model. The algorithm assumes a non-linear relationship between reward and contextual information, which is more in line with the actual urban environment. The nonlinear contextual bandit robust for data offloading (NCB-rDO) algorithm based on the kernel

UCB algorithm is proposed. The algorithm assumes a non-linear relationship between the reward obtained from offloading data and the dynamic heterogeneity of the helper nodes in the fog network. In addition to this, a discount factor for historical rewards is introduced to improve the kernel UCB algorithm, taking into account sudden changes in the urban environment. Our main contributions are summarized below.

- The CB model for recommender systems is applied to the data offloading problem for UWSNs. To more closely match the actual urban environment, the problem of offloading data under a non-stationary model is considered, i.e., the reward distribution is not fixed.
- The architecture based on the fog network makes effective use of the heterogeneity of the nodes in the network. Using the heterogeneous state information of nodes as contextual information, the feature vector based on the heterogeneous information enables nodes to collaborate to offload data.
- The scheme we propose is not a direct application of the CB model and therefore the traditional analysis is not applicable. Experimental results show that the CB-based data offloading strategy is robust to dynamic urban environments.

2 Related Works

In recent years, data offloading has emerged as a promising technology for the efficient use of distributed computing resources, and plenty of researches have been carried out. For example, Li et al. [12] proposed an optimal offloading policy for heterogeneous end-users under buffer constraints. The authors define the objective of achieving maximum mobile data offloading as multiple linear constraint maximization problems with finite storage and propose various algorithms to solve this optimization problem for different offloading scenarios. However, they are considering data offload from the infrastructure to the mobile device. The focus of the work in this paper, however, is on offloading data from sensors to the fog infrastructure. This can lead to different patterns of data loss, the former due to adequate infrastructure resources and data loss problems associated with signaling, the latter focusing on problems caused by a lack of resources for sensors, and more on completing data offloading before resources are consumed. The authors of the literature [13] propose a collaborative offloading algorithm that logically divides sensor nodes into in-need nodes and helper nodes through a central network controller, based on a Markov chain model. This paper also aims to achieve reliable data communication through device collaboration, but our solution aims to enable collaboration between stationary and mobile sensors, rather than between mobile gateways and static sensors. It makes sense to study the data loss problem in different application scenarios, as the topology, node locations, and resources of real sensor networks are dynamic. In this paper, multiple hops are required to complete the data offload. Gao et al. [14] attempted to maximize the data collected by the mobile sink, proposing a scheme by designating the nearest node as the intermediate data collector. Their solution assumes that the node has enough storage to store all the data and that the sink mobility is fixed and known in advance. However, in practice, the shareable computing resources of nodes change over time and mobility is unpredictable. Similarly, Wen et al. [15] constructed

an energy-sensing path for the mobile sinks to collect data when the location of the sensors and the mobile receiver were known. There is research focusing on learning the movement patterns of the sink to improve data collection from the sensors. Pozza et al. [16] considered IoT scenarios through a predictive framework based on temporal difference learning but did not address the issue of communication reliability. Our work considers how to guarantee a certain offloading efficiency when the node state changes abruptly.

3 System Model and Problem Formulation

The problem of data offloading can be naturally modeled as a CB problem. Task nodes are treated as players and helper nodes are treated as arms. In each round, the player expects to select the arm with the highest reward. Similarly, during data offloading at each time slot, the task node offloads data to the helper node and expects a high reward, such as a high level of data offloading success rate.

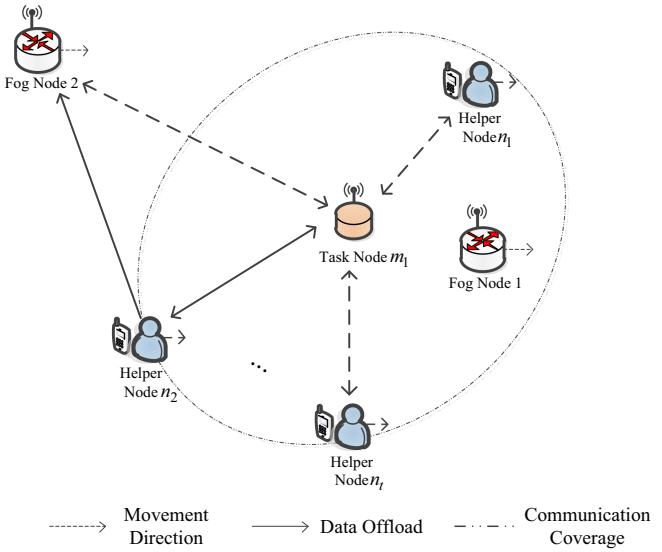

Fig. 1. Cooperative offloading system model

The contextual information-based coordinated offloading scheme for UWSNs is shown in Fig. 1. Suppose there are M task nodes, N $\geq$ M mobile helper nodes, and mobile fog nodes in the model. Time is divided into discrete time slots. T is used to characterize the total number of time slots, and t is used to characterize an arbitrary moment, where $1 \leq t \leq$ T. At each time slot t, the task node broadcasts a data offload request and then selects a helper node based on its observation history. The task node offloads its data to the fog node with the assistance of the helper node. The task node $m \in \{1, \ldots, M\}$ selects a helper node $n \in \{1, \ldots, N\}$, and the task node m gets an

instantaneous reward r_{t,n_t} after the data is offloaded to the fog node. In each period t, the process is as follows: (i) The heterogeneous information $x_{t,n}$ is fed back to the task node as a vector. $x_{t,n}$ is the contextual information of the helper node n at the time slot t. The heterogeneity of helper nodes is various in the same slot. The heterogeneity of the nodes reflects the quality of the nodes, including memory, transmission rate, location, etc. (ii) Selection of helper node: The task node observes heterogeneity information and selects high-quality helper nodes based on the rewards observed in historical trials. (iii) Reward feedback: The fog node feeds back the reward r_{t,n_t} to the task node.

The data offloading problem is constructed as a non-stationary MAB problem, taking into account the dynamics of the urban environment. In this problem, the distribution of the reward can change at an unknown moment. Without loss of generality, we normalize $r_{t,n_t} \in [0, 1]$. Our goal is to achieve a high level of offload success rate, with the reward equal to one when the selected helper node completes the data offload and zero otherwise. The aim is to maximize the above-expected total reward. A discount factor $\tau \in (0, 1)$ was introduced to treat historical rewards with different weights, i.e., a greater weight was given to more recent rewards and a lesser weight to earlier ones.

The mean of the random variable r_{t,n_t} is $\upsilon_{t,n_t} = E[r'_{t,n_t}]$, where r'_{t,n_t} is the average value of discounted rewards. The mean value of the reward distribution on the optimal helper node in each round is $\upsilon_t^{opt} = \max_{n \in N}\{\upsilon_{t,n_t}\}$. In the MAB framework, the regret value is used to measure the performance of a bandit algorithm. The regret value is described as the difference between the total system gain obtained with the best strategy in the ideal case and the total gain obtained with that strategy π. In the time-varying case, the mathematical expression for the regret value of non-stationary bandit after T time slots can be expressed as follow:

$$regret(T) = \sum_{t=1}^{T} \upsilon^{opt} - E[\sum_{t=1}^{T} \upsilon_t^{\pi}] \tag{1}$$

Where $\pi(t) = n$ indicates that, at time slot t, the task node selects the helper node n to offload the data.

4 NCB-rDO Design: Nonlinear Contextual Bandit Robust for Data Offloading

In this section, we consider the case where no prior knowledge of the reward distribution is provided throughout the offloading process, but the distribution is assumed to change over time in all rounds. The NCB-rDO algorithm is proposed, the algorithm has a self-learning capability implemented in a dynamic urban environment, based on the CB model and discounting of rewards.

4.1 Kernel UCB Algorithm

Based on the kernel UCB algorithm, the expected reward and heterogeneous feature information of helper nodes are assumed to be nonlinearly related. In complex urban

wireless sensor networks, it is practical and reasonable to explore the non-linear relationship between the reward of offloading data and the heterogeneity of the nodes. The study [11] shows how to obtain the kernel UCB algorithm. Kernel methods provide a way to extract from observations possibly non-linear relationships between the contexts and the rewards while only using similarity information between contexts. For the nonlinear CB model, we assume that there exists a mapping function $\phi : R^d \rightarrow H$ mapping data to the Hilbert space, for which ϕ, there exists $\theta^* \in H$. Therefore, the expected reward is linearly related to the heterogeneous information of the helper nodes as follows.

$$E\left[r_{t,n_t} \middle| x_{t,n}\right] = \phi(x_{t,n})\theta_n^* \tag{2}$$

According to the mapping function ϕ, the kernel function is defined as $k(x, x^{'}) = \phi(x)^T\phi(x^{'}), \forall x, x^{'} \in R^d$, and the kernel matrix of data $\{x_1, \ldots, x_t\} \in R^d$ is denoted as $K_t = \{k(x_i, x_j)\}_{i,j \leq t}$.

First, the estimated value of the reward $\widehat{\mu}_{t+1,n} = \phi(x_{t+1,n})^T\theta_t$ needs to be predicted, where θ_t is the minimum value of the regularized least-squares function:

$$L(\theta) = \gamma\|\theta\|^2 + \sum_{i=1}^{t-1}(r_i - \phi(x_i)^T\theta)^2 \tag{3}$$

Φ_t is used to denote $[\phi(x_1)^T, \ldots, \phi(x_{t-1})^T]^T$. Let $L(\theta)$ be equal to zero, and the following Eq. (4) can be obtained by taking the partial derivative of θ. The equation satisfies the solution to the problem of minimizing $\theta_t = \min_{\theta \in H} L(\theta)$.

$$(\Phi_t^T\Phi_t + \gamma I)\theta_t = \Phi_t^T y_t \tag{4}$$

where $y_t = \{r_{1,n_1}, \ldots, r_{t,n_t}\}^T$. Equation (4) is rearranged to get $\theta_t = \Phi_t^T\alpha_t$, where $\alpha_t = \gamma^{-1}(y_t - \Phi_t\theta_t) = \gamma^{-1}(y_t - \Phi_t\Phi_t^T\alpha_t)$. The kernel matrix is used to express this equation, i.e., $\alpha_t = (K_t + \gamma I)^{-1}y_t$. We denote $k_{t,x} = \Phi_t\phi(x) = [k(x_1, x), \ldots, k(x_{t-1}, x)]^T$ then Eq. (5) is obtained as:

$$\widehat{\mu}_{t,n} = k_{t,x_{t,n}}^T(K_t + \gamma I)^{-1}y_t \tag{5}$$

While the computation of θ_t would require evaluating $\phi(x_i)$ for every data point x_i, the dualized representation of the prediction (5) allows computing $\widehat{\mu}_{t,n}(x)$ only from the elements of the kernel matrix.

Next, the width of the confidence interval for the predicted value of the reward needs to be calculated. For linear bandits, the appropriate bandwidth is constructed, based on the Mahalanobis distance of $\phi(x_{t,n})$ from the matrix Φ_t:

$$\widehat{\sigma}_{t,n} = \sqrt{\phi(x_{t,n})^T \left(\Phi_t^T \Phi_t + \gamma I\right)^{-1} \phi(x_{t,n})} \tag{6}$$

Since the matrices $(\Phi_t^T \Phi_t + \gamma I)$ and $(\Phi_t \Phi_t^T + \gamma I)$ are strictly positive definite, the following equation holds:

$$(\Phi_t^T \Phi_t + \gamma I)\Phi_t^T = \Phi_t^T (\Phi_t \Phi_t^T + \gamma I) \tag{7}$$

$$\Phi_t^T (\Phi_t \Phi_t^T + \gamma I)^{-1} = (\Phi_t^T \Phi_t + \gamma I)^{-1} \Phi_t^T \tag{8}$$

The Eq. (8) extracts the Mahalanobis distance as

$$(\Phi_t^T \Phi_t + \gamma I)\phi(x) = \Phi_t^T k_{t,x} + \gamma \phi(x) \tag{9}$$

from which it can be inferred that

$$\phi(x) = \Phi_t^T (\Phi_t \Phi_t^T + \gamma I)^{-1} k_{t,x} + \gamma (\Phi_t^T \Phi_t + \gamma I)^{-1} \phi(x) \tag{10}$$

and left multiplied by $\phi(x)$, express $\phi(x)^T \phi(x)$ as

$$k_{t,x}^T (\Phi_t \Phi_t^T + \gamma I)^{-1} k_{t,x} + \gamma \phi(x)^T (\Phi_t^T \Phi_t + \gamma I)^{-1} \phi(x) \tag{11}$$

The above equation is rearranged to give the expression for the width of the confidence interval as

$$\widehat{\sigma}_{t,n} = \gamma^{-1/2} \sqrt{k(x_{t,n}, x_{t,n}) - k_{t,x_{t,n},}^T (K_t + \gamma I)^{-1} k_{t,x_{t,n}}} \tag{12}$$

This equation deals only with the inner product between elements. At each time slot t, the algorithm selects helper nodes n_t satisfying:

$$n_t = \arg\max_{n \in N} \widehat{\mu}_{t,n} + \widehat{\sigma}_{t,n} \tag{13}$$

A pseudo-code description of the kernel UCB algorithm based on data offloading is given below.

Algorithm 1 Kernel UCB With Online Updates [11]

Input and Initialization:

$\mu_0 \leftarrow [1,0,...,0]^T, y_0 \leftarrow \varnothing,\ (K_t)_{ij} = k(x_i, x_j)$

$k_t(x_{t,n}) = [k(x_{1,n_1}, x_{t,n}),...,k(x_{t-1,n_{t-1}}, x_{t,n})]^T$

γ, η regularization and exploration parameters

Run:

for t=1 to T do

Choose $n_t = \arg\max_{n \in N} \mu_n$ and get a reward r_{t,n_t}

Update $y_t = [r_{1,n_1},...,r_{t,n_t}]^T$

if t =1 then

$K_t^{-1} \leftarrow 1 / k_t(x_{t,n}) + \gamma$

else {Online update of the kernel matrix inverse}

$\omega \leftarrow (k_1(x_{1,n_1}),...,k_{t-1}(x_{t-1,n_{t-1}}))^T$

$K_{22} \leftarrow (k(x_{t,n_t}, x_{t,n_t}) + \gamma - \omega^T K_{t-1}^{-1} \omega)^{-1}$, $K_{11} \leftarrow K_{t-1}^{-1} + K_{22} K_{t-1}^{-1} \omega \omega^T K_{t-1}^{-1}$

$K_{12} \leftarrow -K_{22} K_{t-1} \omega$, $K_{21} \leftarrow -K_{22} \omega^T K_{t-1}^{-1}, K_t^{-1} \leftarrow [K_{11}, K_{12}; K_{21}, K_{22}]$

end if

for n =1 to N do

$\sigma_{t,n} \leftarrow \sqrt{k_t(x_{t,n})^T K_t^{-1} k_t(x_{t,n})}$

$\mu_{t,n} \leftarrow (k_t(x_{t,n})^T K_t^{-1} y_t + \frac{\eta}{\gamma^{1/2}} \sigma_{t,n})$

end for

end for

4.2 Discounting the Rewards

In many application areas, temporal changes in the structure of the reward distribution are inherent to the problem. We focus on a discounted MAB formulation that allows for a wide range of temporal uncertainty rewards due to the different demands of the dynamic urban environment. In the presence of uncertainty, players (task nodes) facing a range of decisions need to use information gleaned from past observations when attempting to optimize future actions. Knowing that undetected changes will lead to grossly inaccurate estimates, players need to discount the weights of earlier indicator observations when estimating indicators in an ever-changing environment.

To estimate the instantaneous expected reward, this scheme averages past rewards with a discount factor that gives more weight to the most recent observations. Adaptability to changes in parameters does mean reducing the influence of observations made long in the past, which means using weights that increase with time. At each time slot t, when a helper node is selected, the average value of discounted rewards and the number

of discounts are given by

$$r'_{t,n_t} = \frac{1}{\overline{q}_{t,n_t}} \sum_{s=1}^{t} \tau^{t-s} r_{t,n_t} \mathbb{1}_{\{n_t \in N\}} \tag{14}$$

$$\overline{q}_{t,n_t} = \sum_{s=1}^{t} \tau^{t-s} \mathbb{1}_{\{n_t \in N\}} \tag{15}$$

Where the discount factor $\tau \in (0, 1)$. The proposed policy, referred to as a nonlinear contextual bandit robust for data offloading algorithm, is shown in Algorithm 2.

Algorithm 2 NCB-rDO

//Initialization
for t=1 to N do
  Select helper node $n \in N(t)$ such that $n = (t+1) \bmod N$
    $r'_{t,n_t} = r_{t,n_t}$
    $\overline{q}_{t,n_t} = 1$
end for
// MAIN LOOP
while 1 do
  for t=N+1 to T do
    Run algorithm 1 to select the helper node n_t that maximizes
      $\widehat{\mu}_{t,n} + \widehat{\sigma}_{t,n}$
    Update r'_{t,n_t}, $\overline{q}_{t,n_t}$ accordingly.
end while

5 Simulation

5.1 Evaluation Metrics

The efficiency and performance of the offload scheme are verified by the following performance metrics.

- *Average reward* is defined as the mean of the reward feedback after data offloading to the fog node.
- *The selection count* is defined as the number of corresponding helper nodes selected based on the CB model. It is used to test the robustness of the algorithm when the contextual information of the helper nodes changes abruptly.
- *Success offloading rate* is defined as the rate of success offloading counts compared to overall rounds.

5.2 Methodology and Simulation Setup

The Python software environment was used to evaluate the performance of the proposed data offloading scheme under dynamic conditions according to the scheme proposed in Sect. 4. The hardware environment used for the experiments is Intel(R) Core i5-4210M (2.60 GHz) with 8 GB of RAM. In the proposed data offloading scheme, the dynamic heterogeneity of helper nodes in each round is used as contextual information. The heterogeneity of each helper node is assumed to consist of four components, namely memory margin, transmission speed, residual energy, and movement probability. The memory margins follow a uniform distribution with distribution parameters ranging from 0.04 to 0.06. The residual energy follows an exponential distribution with the distribution parameter $\boldsymbol{\delta_n} = [\delta_n]_{n=1,\ldots,10}$, the range of $\boldsymbol{\delta_n}$ is from 0.1 to 1. The transmission rate of the helper node takes values from 2 to 24 Mbit/s. To evaluate CB-based offloading strategies in a more realistic context, ONE Simulator v1.6.0 was used to generate simulations of urban road-based movement trajectories in which pedestrians carrying smartphones (as helper nodes) walk along a street at a speed of 0.5–1.5 m/s. At the beginning of the simulation, all sensors were initialized with the same probability of movement, and then each sensor dynamically updated the observed probability of encounter with the fog node. There are ten helper nodes and mobile fog nodes in the simulation network, which are placed in an area with a radius of 1000 m. The length of the time slot is set to 20 ms. A high level of successful offloading rate is used as a criterion for selecting the helper node. The helper node is selected to enable the collaborative offloading system to obtain a satisfactory offloading ratio in the CB framework. r'_{t,n_t} is denoted to the average value of the discounted reward taken by the helper node n at the time slot t. To simplify the process, we assume that follows a Gaussian distribution with parameters r'_{t,n_t}.

The NCB-rDO comes with two hyperparameters including gamma and eta. To optimize these hyperparameters we propose to perform a grid search as in Fig. 2, where the value of gamma will be between zero and three, excluding zero, and the value of eta will be between zero and one. The results will be displayed on the grid with color bars. The experimental results show that the mean reward reaches a maximum of 0.77 when gamma takes the value of 0.5 and eta takes the value of 0.3.

5.3 Simulation Results

The proposed collaborative offload policy is compared with three other bandit options. LinUCB [10] is the classical contextual bandit algorithm, which assumes a linear relationship between the expected reward of the selected helper node and the contextual information. The UCB [17] algorithm considers the average reward and confidence interval of the arms. Each helper node corresponds to a confidence interval and the helper node with the largest upper confidence interval is selected for data offloading. Epsilon-Greedy algorithm: the relationship between random numbers and a priori epsilon values is used to decide whether to select the helper node with the highest reward or to select the helper node at random.

Figure 3 illustrates the change in the success rate of data offloading as the number of trial rounds increases for the four considered offloading schemes. The NCB-rDO algorithm could achieve an offloading success rate of 80.28%; the LinUCB algorithm

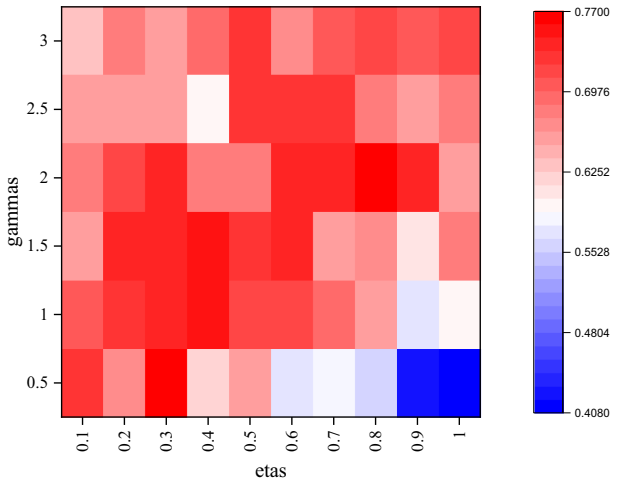

Fig. 2. Grid search average reward value for NCB-rDO

could achieve an offloading success rate of 74.67%; Compared to the first two contextual bandit algorithms, the UCB algorithm and the Epsilon-Greedy algorithm have a lower offloading success rate of 65.83% and 56.06%, respectively. The Epsilon-Greedy algorithm rigidly divides the selection process into an exploration phase and an exploitation phase. The exploration is performed with the same prior probability for all helper nodes and does not make use of any historical information, including the number of times a helper node has been selected and the probability of a helper node receiving a reward. As a result, the algorithm has the lowest success rate in offloading data.

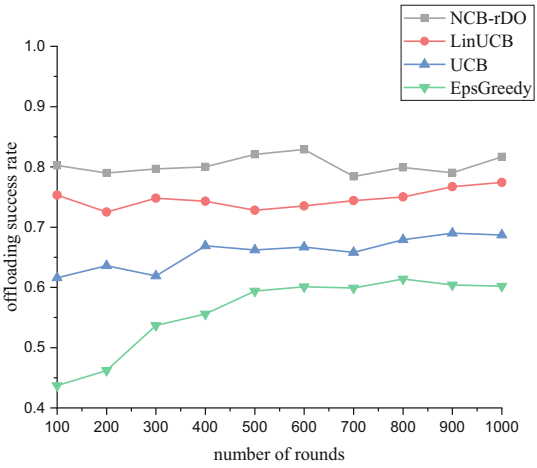

Fig. 3. Offloading success rate for four data offloading policies.

Figure 4 illustrates the change in the average reward for the four data offloading scenarios as the number of experimental rounds increases. High-level success offloading rate is linked with contextual bandit rewards. To obtain a higher reward, the gamma in the NCB-rDO algorithm was set to 0.5 and the eta was set to 0.3. The average reward for NCB-rDO is higher when compared to the other three schemes. Compared to LinUCB, NCB-rDO assumes the expected reward and heterogeneity of helper nodes to be non-linear, which is more in line with realistic scenarios.

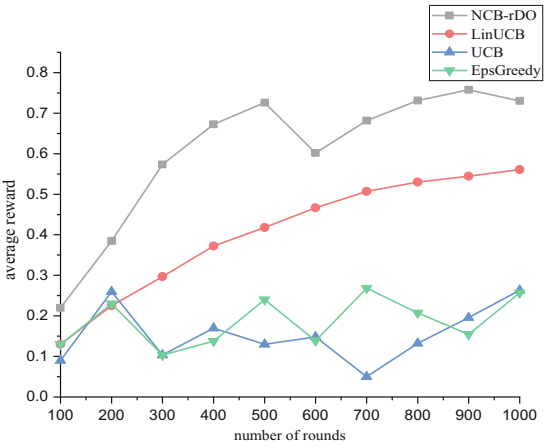

Fig. 4. The average reward for four data offloading policies.

Table 1. Time cost

	NCB-rDO	LinUCB
Feature vectors feedback	0.0227 s	0.0174 s
Decision and Reward feedback	2.9731 s	1.1271 s
Total	2.9958 s	1.1445 s

Table 1 shows a comparison of the two contextual bandit algorithms, the NCB-rDO and LinUCB algorithms, in terms of execution time. It consists of three parts of time: contextual information feedback, helper node decision, and reward feedback. The values are derived from 1000 iterations. The total time cost of LinUCB is 1.1445 s. By contrast, the high offloading success rate and average reward policy—NCB-rDO—costs a longer time of approximately 2.9958 s.

The robustness of the LinUCB algorithm and the NCB-rDO algorithm was tested in a highly dynamic environment. The state information for node 2 is assumed to be initially good, i.e., large memory margin, fast transmission speed, etc., while the state information for node 5 is initially poor. Figure 5 shows that the number of selections and the node state information are positively correlated. When $t = 700$, helper nodes 2 and

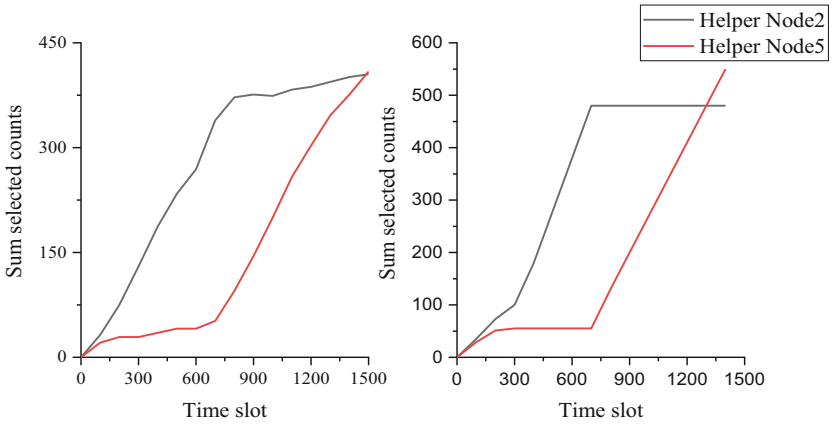

Fig. 5. Robustness performance for Lin UCB and NCB-rDO

5 suffer an abrupt change, i.e., the state information of node 2 is poor, while that of node 5 is good. As shown in Fig. 5, the NCB-rDO algorithm is more sensitive to changes in contextual information compared to the LinUCB algorithm. When $t > 700$, the number of selections for both nodes changes more significantly. Both the LinUCB and NCB-rDO algorithms are based on the CB model, with the difference that the former assumes a linear relationship between contextual information and reward, while the latter assumes a non-linear relationship. The non-linear relationship is more consistent with the actual mutating environment, and NCB-rDO can show stronger robustness than LinUCB.

6 Conclusion

A new collaborative data offloading algorithm based on the CB model is proposed without prior knowledge of the state of the nodes. The framework based on fog computing makes more rational use of the heterogeneity of the nodes, thus avoiding data loss problems and increasing the success rate of data transmission. Also, based on a non-stationary bandit model, the idea of discounted rewards is introduced to weigh different historical rewards. The proposed scenario allows the best helper node selection only with a bit of contextual information about nodes. Numerical analysis verifies that the NCB-rDO algorithm avoids data loss problems in urban environments and also verifies the robustness of the algorithm in the face of abrupt changes. The multi-source data offloading problem is not a simple overlay of the single-source data offloading problem. Therefore, our next work will consider a coordinated multi-source data offloading scheme based on a contextual bandit model.

References

1. Miorandi, D., Sicari, S., Pellegrini, F.D., Chlamtac, I.: Internet of things: vision, applications and research challenges. Ad Hoc Netw. **10**(7), 1497–1516 (2012)

2. Chiang, M., Zhang, T.: Fog and IoT: an overview of research opportunities. IEEE IoT J. **3**(6), 854–864 (2016)
3. Yuan, D., Kanhere, S.S., Hollick, M.: Instrumenting wireless sensor networks—a survey on the metrics that matter. Pervasive Mob. Comput. **37**, 45–62 (2017)
4. Yu, T., Wang, X., Shami, A.: A novel fog computing enabled temporal data reduction scheme in IoT systems. In: The 2017 IEEE Global Communications Conference, pp. 1–5, December 2017
5. Bonomi, F., Milito, R., Natarajan, P., Zhu, J.: Fog computing: a platform for internet of things and analytics. In: Bessis, N., Dobre, C. (eds.) Big Data and Internet of Things: A Roadmap for Smart Environments. SCI, vol. 546, pp. 169–186. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-05029-4_7
6. Chiang, M., Zhang, T.: Fog and IoT: an overview of research opportunities. IEEE Internet Things J. **3**(6), 854–864 (2016)
7. Wen, Z., Yang, R., Garraghan, P., Lin, T., Xu, J., Rovatsos, M.: Fog orchestration for internet of things services. IEEE Internet Comput. **21**, 16–24 (2017)
8. Zhou, P., Jiang, T.: Toward optimal adaptive wireless communications in unknown environments. IEEE Trans. Wirel. Commun. **15**(5), 3655–3667 (2016)
9. Maghsudi, S., Hossain, E.: Multi-armed bandits with application to 5G small cells. IEEE Wirel. Commun. **23**(3), 64–73 (2016)
10. Li, L., Chu, W., Langford, J., Schapire, R.E.: A contextual-bandit approach to personalized news article recommendation. Presented at the Proceedings of the 19th International Conference on World Wide Web, Raleigh, North Carolina, USA (2010)
11. Valko, M., Korda, N., Munos, R., Flaounas, I., Cristianini, N.: Finitetime analysis of kernelised contextual bandits. In: Proceedings of Uncertainty Artificial Intelligence, pp. 654–663 (2013)
12. Li, Y., Qian, M., Jin, D., Hui, P., Wang, Z., Chen, S.: Multiple mobile data offloading through disruption tolerant networks. IEEE Trans. Mob. Comput. **13**, 1579–1596 (2014)
13. Kortoçi, P., Zheng, L., Joe-wong, C., Francesco, M.D., Chiang, M.: Fog-based data offloading in urban IoT scenarios. In: IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, pp. 784–792 (2019)
14. Gao, S., Zhang, H., Das, S.K.: Efficient data collection in wireless sensor networks with path-constrained mobile sinks. IEEE Trans. Mob. Comput. **10**(4), 592–608 (2011)
15. Wen, W., Zhao, S., Shang, C., Chang, C.-Y.: EAPC: energy-aware path construction for data collection using mobile sink in wireless sensor networks. IEEE Sens. J. **18**(2), 890–901 (2018)
16. Pozza, R., Nati, M., Georgoulas, S., Gluhak, A., Moessner, K., Krco, S.: CARD: context-aware resource discovery for mobile internet of things scenarios. In: IEEE WoWMoM 2014, pp. 1–10, June 2014
17. Agrawal, R.: Sample mean based index policies by O (log n) regret for the multi-armed bandit problem. Adv. Appl. Probab. **27**, 1054–1078 (1995)