Toward Multi-sided Fairness: A Fairness-Aware Order Dispatch System for Instant Delivery Service

Zouying Cao[1], Lin Jiang[2], Xiaolei Zhou[3(✉)], Shilin Zhu[1], Hai Wang[1], and Shuai Wang[1]

[1] School of Computer Science and Technology, Southeast University, Nanjing, China
{zouyingcao,shilinzhu,hai,shuaiwang}@seu.edu.cn
[2] School of Cyber Science and Engineering, Southeast University, Nanjing, China
linjiang@seu.edu.cn
[3] The Sixty-Third Research Institute, National University of Defense Technology, Changsha, China
zhouxiaolei@nudt.edu.cn

Abstract. Instant delivery platforms, equipped with professional couriers to provide convenient delivery services, have emerged rapidly in many cities. For the benefit of platforms, many researchers focus more on maximizing overall efficiency but ignore individual fairness. Current fairness research in mobile systems mainly concentrates on one-sided or two-sided relationships, such as drivers and customers. However, instant delivery services have two new characteristics in fairness: (i) **multi-stakeholder involvement**, namely couriers, merchants and users should be considered comprehensively; (ii) more complicated matching relationship because of **the concurrent dispatch mode**, meaning one courier will handle multiple orders simultaneously. To handle this multi-sided fairness problem, our paper proposes a novel order dispatch system to balance the platform revenue and multi-stakeholder fairness. Motivated by the analysis of real-world datasets, we formulate the order dispatch problem as a sequential decision-making problem and incorporate multi-sided fairness into the decision criteria. Then, we design a multi-sided fairness-aware deep reinforcement learning algorithm to solve large-scale decision problem, with the fairness relying on Least Misery Fairness definition for users and Variance Fairness definition for couriers and merchants. Finally, extensive experiments show the effectiveness of our model in balancing multi-sided fairness among stakeholders and long-term profits of the whole platform.

Keywords: Order dispatch · Multi-sided fairness · Instant delivery · Reinforcement learning

1 Introduction

With the rapid development of O2O (online to offline) and New Retailing, instant delivery services have gained much popularity and facilitate people's daily lives

L. Wang et al. (Eds.): WASA 2022, LNCS 13472, pp. 303–316, 2022.
https://doi.org/10.1007/978-3-031-19214-2_25

enormously. Driven by the growing demand, many popular platforms (e.g., DoorDash, UberEats, Instacart and MeiTuan) provide fast delivery services to help their customers acquire food, medicine, and groceries quickly. In 2020, mainland China had over 22.85 billion instant delivery orders, marking a year-on-year increase of 25%. Although the rapid growth of instant delivery generates huge economical profit, it leads to multiple challenges in social governance which deserves to be studied to solve.

Most of the research efforts [7,20,21] for instant delivery services concentrate on maximizing the efficiency for order dispatch and improving the experiences at the users' side. However, the issue of fairness in sharing economy attracts widely attentions from the whole society. As the number of registered couriers grows, it is crucial to guarantee the fairness among their incomes, and the same is true for merchants. Besides, violating user fairness is not only ethically fraught, but also unfavorable for securing the platform performance in the long term.

The current fair matching mechanism has several drawbacks. Firstly, some matching models [3,6,9] only pay attention to one-sided fairness but ignore the overall fairness among other stakeholders. Secondly, instant delivery service has concurrent dispatch mode, which means one courier can process multiple orders simultaneously. However, most algorithms only focus on the sequence dispatch mode which is common on ride-hailing platforms [14,19]. Therefore, these latest algorithms are not suitable for instant delivery platforms, and we are interested in investigating a novel order dispatch system to ensure the multi-sided fairness in instant delivery platforms.

Since order dispatch decisions are ordered by time, we can explore the use of Reinforcement Learning (RL) [8,10,22] in the instant delivery serving multi-sided fairness. In addition, with massive historical dispatch and route records, we can amortize equality in multi-sided systems over longer periods (e.g., weeks or months) and then extend the concept of fairness to multiple stakeholders based on an empirical data analysis. Similar ideas of fairness amortization [14] have been utilized in the context of ranking [2,13] and recommender systems [1,12].

However, seeking an optimal fairness-aware order dispatch algorithm applying to this new multi-stakeholder commercial platform is not an easy task due to two challenges: (i) **uncertain multi-sided fairness notions** as different stakeholders may have different perceptions of fairness; (ii) **potential conflicting relationships** within the same stakeholder group and between different stakeholder groups. For example, reducing income inequality among couriers may result in inefficient service as well as loss and disparity in customer utilities.

To tackle the above challenges, we propose an Advantage Actor-Critic-based deep reinforcement learning approach to learn the Multi-sided Fairness-aware order dispatch policy called **A2CMF**. Then, we establish two notions of fairness based on the "variance" fairness semantics for the couriers and merchants to maintain equality, and utilize the "least misery" [18] to guarantee the user waiting time within reasonable bounds. Specifically, we design a policy network in A2CMF which integrates state & action embedding features and two fairness metrics into an accumulated reward. And different from traditional actor-critic algorithm, the action space as input is designed variable to handle the uncertain number of optional couriers considering user fairness constraints.

In summary, the salient contributions of this paper are as follows.

- To the best of our knowledge, we perform the first work on multi-sided (tripartite or more) fairness-aware order dispatch policy in an instant delivery platform. Our approach is proposed with 1,159,371 order records in one month relevant to 595 merchants and about 4,000 couriers. We believe our proposal would contribute to further explore the fair matching issue when the new commercial pattern brings the complicated multiple service suppliers model.
- We consider multi-sided notions of fairness which not only relate to the fair income distribution for couriers, the fair service experience among customers and merchants but also the long-term profitability of multi-stakeholder platforms. Such an idea would be helpful to address the fairness concerned issues in similar sharing economy scenarios.
- More importantly, to better train the A2CMF Network, we design a data-driven simulator to model the real-time instant delivery environment with dynamic demand & supply, spatial-temporal features and complicated courier behaviors. Then we evaluate the performance of A2CMF through extensive experimentation with data from Eleme (one of the largest instant delivery companies in China). The evaluation results show that our A2CMF achieves a 21.3% increase in total revenue, improves the profit fairness of couriers by 9.7%, and reduces customers' waiting time and the benefit gap among merchants by 6.9% and 6.2%, simultaneously.

2 Background and Motivation

2.1 Instant Delivery Scenario

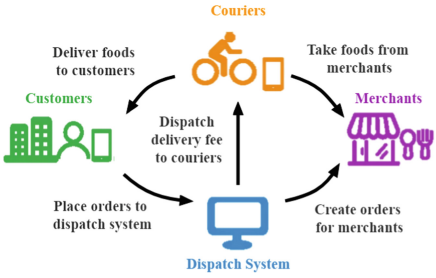

Fig. 1. Four stakeholders in instant delivery

Table 1. Order progress record

Field	Value
User/Courier/Merchant ID	U001/C001/M001
Food Amt/Delivery Fee	32.99/3.8
Promise Delivery Time	3300
Merchant Location	121.45916,31.25554
User Location	121.46889,31.25317
Order Create Time	2020/10/1 11:59:00
Accept Order Time	2020/10/1 12:03:00
Arrive Restaurant Time	2020/10/1 12:07:00
Pickup Time	2020/10/1 12:09:00
Delivery Time	2020/10/1 12:17:00

As illustrated in Fig. 1, a typical instant delivery service involves four stakeholders: **couriers**, **merchants**, **customers** and the **platform**. And their corresponding information combined with five critical timestamps will be recorded during the order dispatching process (e.g., listed in Table 1). Then, the roles of the four stakeholders in instant delivery will be briefly introduced.

(i) **Couriers** are assigned order tasks by the platform, and need to pick up orders in merchants and deliver to customers in time. The fairness appeal of couriers is that they can get the same labour efficiency when having the same working hours. (ii) **Merchants** receive orders from customers and are arranged couriers by platform. From the perspective of merchants, they want to get couriers to pick up prepared orders as soon as possible. (iii) When **customers** place orders through the platform and look forward to acquiring them on schedule, it's better to have early delivery. (iv) As the principal of dispatch algorithm, **platform** has the primary aim to obtain more benefit, but it also has the responsibility to consider the fair requirements of the other three stakeholders. Only achieve a trade-off among the above four aspects, can the instant delivery platform realize a stable operation in the long run.

2.2 Characteristics of Fairness in Order Dispatch

Given the historical delivery order data, we conduct a data-driven order dispatch pattern analysis and obtain the following observations:

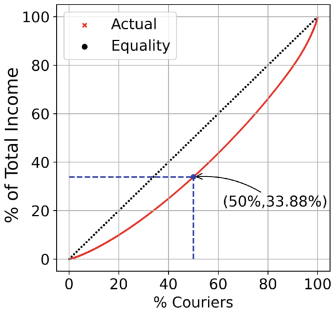

Fig. 2. Lorenz curve of courier income

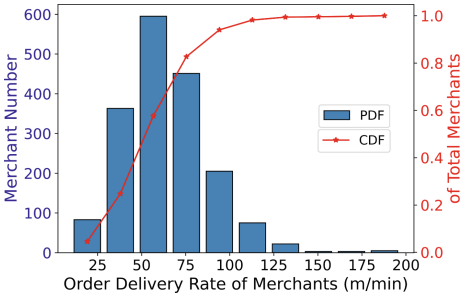

Fig. 3. Merchant fairness index

1. **Income Inequality among Couriers.** Figure 2 shows that after one day, 50% couriers only earned 33% of total income, while 20% most successful couriers get 35%. Couriers in the bottom ten percent of income almost made little money, which represents unequal income distribution among couriers.
2. **Unfair User Experience.** From user comments on the platform, we observe that positive and negative comments coexist and the comments are even polarized, showing unfair service experience issues among customers.
3. **Inequality in Merchant Benefit.** We further analyze the order delivery rate of each merchant and find the inequality in merchant benefit. It is demonstrated in Fig. 3 that 20% of merchants are severely lower than the average level while another 20% are significantly higher than the average.

3 System and Formulation

3.1 System Overview

We present the overview of our system design in Fig. 4 which is composed of **three** modules.

1. **An Environment Simulator for Instant Delivery.** We introduce a simulator design that models the events of order generation, order assignments and key stakeholder behaviors such as distributions across the city, along with changes in weather and traffic conditions in the real world.
2. **A State & Action Feature Extraction Module.** This part serves as a feature extractor to characterize multiple attributes important for order dispatching decisions, including order features, spatial features, temporal features and environmental features.
3. **A2CMF-Dispatch Model.** This model aims to learn an optimal fairness-aware order dispatch policy by calculating the long-term value for each candidate dispatch action via the Actor Network and then achieves a more stable and efficient model learning process via the Critic Network.

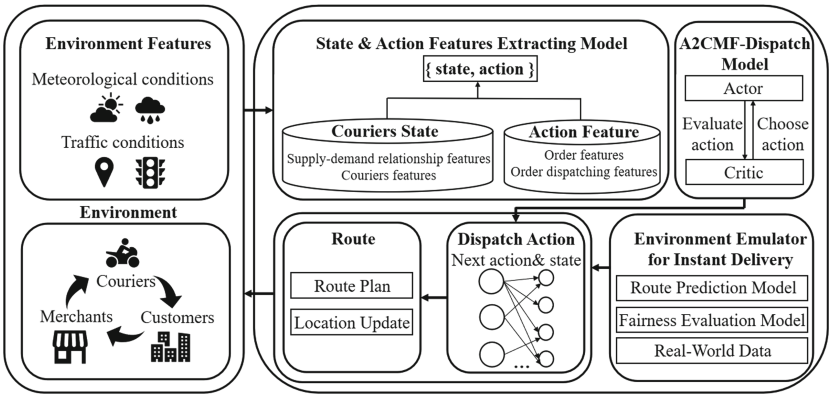

Fig. 4. System architecture of A2CMF

3.2 Problem Formulation

In this paper, we formulate the multi-sided fairness-aware order dispatch problem as *a multi−agent Markov decision process*, which is characterized by five components: $\{\mathcal{S}, \mathcal{A}, \mathcal{R}, \gamma, \pi\}$, i.e., the integration of states $\mathcal{S}$, the courier action space $\mathcal{A}$, the multi-sided fairness-oriented reward $\mathcal{R}$, a discount factor γ and the policy π to make fair matching decisions. Formally, we present the training process as finding a target policy $\pi(\mathrm{a}|\mathrm{s})$ so that dispatching actions τ according to $\pi(\mathrm{a}|\mathrm{s})$, would lead to the maximum expected cumulative reward:

$$max_{\pi_\theta} \mathbb{E}_{\tau \sim \pi_\theta}[R(\tau)], \tag{1}$$

where $R(\tau) = \sum_{t=0}^{|\tau|} r(s_t, a_t)$ and θ denotes policy parameter. The specific definitions of the *multi−agent Markov decision process* in our **A2CMF** are listed below and policy π is introduced detailed in Sect. 4.2.

- **Agent Set.** We consider each candidate courier as an agent, and all couriers share the same dispatch strategy. In our system, the dispatch strategy is under centralized training, but does a decentralised execution based on every individual agent (courier) [16].
- **State** $\mathcal{S}$. We divide one day into $\mathcal{T}$ time slots and represent the state at time slot t as $s_{t\sim\mathcal{T}} = \{\mathcal{P}_t, \mathcal{ST}_t, \mathcal{D}_t, \mathcal{C}_t\}$, where $\mathcal{P}$ is the personal attribute of courier, $\mathcal{ST}$ is the set of spatiotemporal features, $\mathcal{D}$ is the global information about the distribution of stakeholders, and $\mathcal{C}$ are some contextual features.
  - $\mathcal{P}_t$: The personal state of a courier is defined as $\mathcal{P}_t = [loc, n_o, t_o, f, route_p]$, where loc is courier real-time location, n_o is the number of existing orders, t_o is his/her on-duty time and f marks whether this courier can deliver the order without disturbing the customer fairness index. Last, we use the idea designed by Zhang et al. [21] to predict $route_p$.
  - $\mathcal{ST}_t$: Note that couriers' locations are continuously changing, which will affect future order receive rate. We define a local-view state $\mathcal{ST}_t = \frac{N_o}{N_c}$ capturing the income opportunity where N_o and N_c are the total number of orders and rival couriers along the planning route, respectively.
  - $\mathcal{D}_t$: Shared by all couriers, the global-view state $\mathcal{D}$(consists of $\mathcal{O}, \mathcal{CO}, \mathcal{M}$ and $\mathcal{U}$) depicts the demand and supply distribution. The four matrices record the online number of four parts(i.e., orders, couriers, merchants, users) in each grid, representing a fine-grained distribution.
  - $\mathcal{C}_t$: In instant delivery, customers are tolerant of delayed packages [21] because of factors such as bad weather, rush hours and so on. So, we take into account those contextual information via one-hot encoding.
- **Action** $\mathcal{A}$. The agent action in our proposal is to recommend the optimal order to the courier waiting to be allocated. Thus, action features involve: (1) **order features** including price p, the create time t_c, the merchant location l_m and the customer location l_c; (2) **order dispatching attributes** including the delivery distance, the increased route time if the courier takes this order. Eventually, after choosing an optimal courier to take action, all related states need a proper transition $P(s_{t+1}|s_t, a_t) : S \times A \to S$.
- **Reward** $\mathcal{R}$. As the reward function, $\mathcal{R} = r(s_t, a_t)$ denotes the immediate reward of the action a_t at specific state s_t. It is designed to reach a balance between the overall platform revenue and the fairness among stakeholders:

$$r^{(i)}(s_t, a_t) = (1 - \alpha - \beta)PE(i, t) + \alpha \cdot (-CF(t)) + \beta \cdot (-MF(t)) \quad (2)$$

where $\alpha, \beta \in [0, 1]$ balance the profit efficiency and two-sided fairness.

- $PE(i, t)$ is the profit efficiency of the order i in the time slot t and set as

$$PE(i, t) = \gamma^{\Delta t} \times fee_i \quad (3)$$

where Δt is the actual delivery time, γ is the discount factor in regard to the influence of time cost and fee_i is the delivery fee of order i.

- $CF(t)$ is a metric of profit fairness among couriers. For Variance fairness semantic, the fairness index is formulated as:

$$CF(t) = \frac{1}{N_c}\sum_{k=1}^{N_c}(CE(k,t) - \overline{CE(t)})^2, \tag{4}$$

the variance of profit efficiency CE of all N_c online couriers. And CE consists of two components: profit efficiency PE of each $order^i$ delivered by the $courier^k$ and working hours T_{work} measured in one time slot.

$$CE(k,t) = \frac{\sum_{i=1}^{m} PE(i,t)}{T_{work}(k,t)}, \overline{CE(t)} = \frac{1}{N_c}\sum_{k=1}^{N_c} CE(k,t) \tag{5}$$

- $MF(t)$ describes the profit fairness among merchants and borrows idea from Meituan that they think it is fair when merchants' products can be picked up and delivered to customers in time. Based on this, we define MF as a variance of the mean product value PV of all N_m merchants.

$$MF(t) = \frac{1}{N_m}\sum_{m=1}^{N_m}(PV(m,t) - \overline{PV(t)})^2, PV(m,t) = \frac{1}{N_o}\sum_{i=1}^{N_o}\frac{dist}{T_d^{(i)} - T_m^{(i)}} \tag{6}$$

where $N_o, dist, T_m^{(i)}, T_d^{(i)}$ denote the number of orders produced, the delivery distance, the time when $order^i$ is ready and delivered to the user.

Therefore, the Eq. 2 can be converted to Eq. 7.

$$r^{(i)}(s_t, a_t) = (1-\alpha-\beta)\cdot(\gamma^{\Delta t}\cdot fee_i)$$
$$+\alpha\cdot(-\frac{1}{N_c}\sum_{k=1}^{N_c}(CE(k,t) - \overline{CE(t)})^2) + \beta\cdot(-\frac{1}{N_m}\sum_{m=1}^{N_m}(PV(m,t) - \overline{PV(t)})^2) \tag{7}$$

– **Discount factor** γ. γ selected from [0,1] discusses the time-based penalization for the rewards agent achieved in the past, present, and future.

4 Order Dispatch Model Design

In this section, we show how we solve the above formulated fairness-aware order dispatch problem with our advantage actor-critic(A2C)-based model **A2CMF**.

4.1 Environment Simulator Design

As real-world features give crucial content about dispatch decisions and actions can also affect the environment, an environment simulator for instant delivery plays a functional role in the performance of **A2CMF**. To simulate the real order dispatch environment better, we include the following features, generally can be classified into **five** categories:

- **Order Features.** Order features provide the basic information(e.g., price, create time, the corresponding merchant, and customer location).
- **Couriers Features.** Couriers distinguish from each other by their positions, capacity, working hours, number of existing orders, and route planning.
- **Supply-demand Relationship Features.** By capturing the real-time distribution of couriers and orders at the grid level, this kind of features describe the fine-grained supply-demand relationship in the instant delivery platform.
- **Order Dispatching Features (i.e., Action Features).** An order dispatching action is depicted by the planned route, the distance between merchant and courier, and the increased delivery time when the order is added.
- **Environmental Features.** Like meteorological conditions and traffic fleet, environmental features give contextual content about dispatch decisions.

4.2 Advantage Actor-Critic Network

The basic idea of A2C algorithm is that there are two networks, a policy network(i.e., Actor, utilized to calculate the possible long-term reward of the courier-order matching and learn a policy) and a value network (i.e., Critic, a state-value function and leveraged to evaluate the performance of the actor).

In our problem, we collect nearby couriers under customer fairness constraints when a new order is created and extract the pair $\langle state, action \rangle$ as the input of Actor. After feeding them into feature embedding layers respectively, we concatenate two features and feed the result vectors into hidden layers to calculate the long-term matching reward Q. Finally, given all possible $courier^k$-$order^i$ matching value $Q(s_t, a_t)$, policy π is parameterized as

$$\pi(a_t^{k=c}|s_t^i;\theta) = \frac{exp(Q(s_t^i, a_t^{k=c}))}{\sum_{c'=c_1}^{C} exp(Q(s_t^i, a_t^{k=c'}))} \tag{8}$$

where $a_t^{k=c}$ means dispatching $courier^c$ to deliver $order^i$ and θ is weight of Actor.

The second network called Critic judges whether the action selected by policy π is optimal or not and predicts the state-value function defined in Eq. 9.

$$V(s_t;w) = E[\sum_{k=0}^{\infty} \gamma^k r_{t+k}|s_t] \tag{9}$$

where w denote the parameters of the Critic.

Then, Actor are updated in the direction of $\nabla_\theta log\pi(a_t|s_t;\theta)A(a_t, s_t)$ where $A(a_t, s_t)$ is an advantage function ($k = 1$ in our experiment) which estimates the relative benefit of taking action a_t in state s_t and computed as Eq. 10.

$$A(a_t, s_t) = \sum_{i=0}^{k-1} \gamma^i r_{t+i} + \gamma^k V(s_{t+k}) - V(s_t) \tag{10}$$

We update the parameters w of the value function $V(s;w)$ by minimizing the square loss of actual state value and estimated state value:

$$\arg\min_w \frac{1}{2}[r_{t+1} + \gamma V(s_{t+1};w) - V(s_t;w)]^2 \tag{11}$$

4.3 Order Dispatch Based on A2CMF Model

Lastly, the training process of the **A2CMF** model will be introduced in detail.

1. **Get order information in one small period.** At each period, environment generate orders from real datasets containing information such as merchant location, customer location, price, promising delivery time and so on.
2. **Determine the dispatch range.** For each order, couriers in nearby areas have the chance to take this order. Our A2CMF model selects a proper number of candidate couriers within severe constraints of user satisfaction to avoid improper matches disturbing user fairness.
3. **Extract order and pending couriers' features.** After determining the pending couriers, using the environment simulator and feature extraction module, we extract features including spatial-temporal information, route planning information, existing order information and weather information.
4. **Find optimum courier and dispatch order to him/her.** The model receives each order-courier feature and sends them into the A2C network. Then Actor network calculates the order-courier matching value and recommends the optimum courier with Softmax function and Critic network works for advantage function by receiving reward and generating state value.
5. **Simulate and execute couriers' route plan.** After all orders in this small period have been arranged optimum couriers, our system would update the couriers' future route plans based on the environment simulator.
6. **Record feedback reward and update the A2C network.** Our system would record the reward from environment upon couriers finishing one order. And using the reward and state and action information, we can optimize the A2C network making its decisions closer and closer to the final fairness goals.

In a word, based on the environment simulator and feature extraction module, the A2CMF model can reasonably simulate the operation of couriers' movement and order dispatch in instant delivery. Meanwhile, the A2C network and reward based on multi-sided fairness can effectively guide the dispatch system to make a reasonable trade-off between system efficiency and multi-sided fairness.

5 Evaluation

5.1 Evaluation Methodology

Parameter Setting. We implement A2CMF and consider order dispatch in a map of 10×10 spatial grids with 167 time steps (i.e., 5 min as a time slot). At each time step, orders can only be dispatched to the courier whose customers don't wait over 8 min in peak hours and 5 min otherwise. And to guarantee the convergence, we set $\alpha = \beta = 0.3$ to balance the profit and fairness.

Baselines. To show the effectiveness of our system, we compare A2CMF with

- **GT(the ground truth)** is the order dispatch strategy extracted from the data simulated by the simulator in Eleme;
- **RD(random dispatch)** is the algorithm which always selects the courier with random strategy without considering muti-sided fairness;
- **SD2** is the shortest distance based dispatching method [11]. When one new order is created, it will be dispatched to the nearest courier in line with the customer is always right philosophy;
- **IDT** takes into account the influence of the new order added to a courier's route plan, using the increased delivery time based policy;
- **DDQN-as** utilizes a Double-DQN network to learn a order dispatch method in ride-sharing, with additional capability of carrying out action search [17];
- **XgD** is a Xgboost-based dispatch method in instant delivery [21]. Xgboost does ranking considering couriers' income, delivery distance and the increased journey time, and orders are dispatched to the courier ranking first.

Metrics. The evaluation metrics for capturing the fairness and efficiency are:

- **Total revenue (R_p):** From the platform's perspective, we investigate the efficiency of different order dispatch algorithms which is defined as the sum of each order's profit efficiency (Eq. 3).
- **Courier-side profit fairness ($Gini_c$):** We investigate income distributions among n couriers which is given by Gini Coefficient. The lower $Gini_c = \frac{\sum_{i=1}^{n}(2i-n-1)CE_i}{n\sum_{i=1}^{n} CE_i}$ (CE defined in Eq. 5), the better is courier fairness.
- **Merchant-side benefit gap (G_m):** To capture minimum benefit gap guarantee for all merchants, we compute the variance G_m of the merchants' mean product value (Eq. 6). The lower G_m, the smaller is merchant benefit gap.
- **Customer-side metrics:**
  - **Mean average waiting time (M_w):** Although A2CMF ensures Least Misery Fairness guarantee for customers, here we capture how effectively this reduces waiting time M_w on average in comparison to the baselines.
  - **Disparity in waiting time (D_w):** We also calculate the standard deviation of customer waiting time, that is, $D_w = \sqrt{\frac{1}{N_m}\sum_{k=1}^{N_m}(T(k) - M_w)^2}$. The lower the D_w, lesser is the disparity in waiting time.

5.2 Main Performance

Table 2 reports the overall results of our A2CMF model and all the compared baselines concerning our four metrics. As can be seen, A2CMF achieves the most well-rounded performance among all the baselines.

Table 2. Performance comparison

Method	R_p	$Gini_c$	G_m	M_w	D_w
GT	100%	100%	100%	100%	100%
RD	40.2%	66.4%	85.4%	–	–
SD2 [11]	67.4%	131.5%	130.4%	–	–
IDT	92.3%	78.2%	92.8%	81.6%	77.6%
DDQN-as [17]	113.1%	56.3%	111.4%	95.0%	82.3%
XgD [21]	42.8%	76.1%	119.4%	–	–
A2CMF	**137.2%**	**50.8%**	**80.1%**	**93.1%**	**81.8%**

Specifically, our method increases 21.3% of total revenue than DDQN-as which has the second-best performance. Figure 5 gives a visual confirmation that the performance is better in enhancing the total revenue when we choose A2CMF.

In Fig. 6, A2CMF outperforms other baselines in reducing the benefits gap among merchants, with a 19.9% decrease compared with GT. From Fig. 7 and Fig. 8, A2CMF's smallest radian of the Lorenz and the smallest $Gini_c$ illustrate its performance in helping couriers achieve more equitable income distribution.

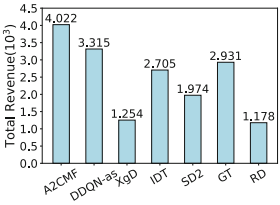

Fig. 5. Total revenue

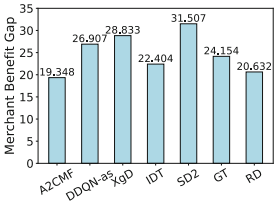

Fig. 6. Benefit gap G_m

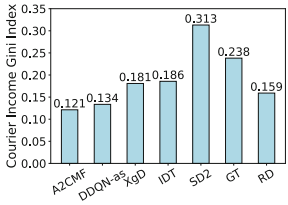

Fig. 7. Courier income $Gini$

In addition, we present the comparison in terms of customer-side metrics with GT, IDT, and DDQN-as. Figure 9 and Fig. 10 show that through our dispatch algorithm, we help users save 6.9% customer's mean waiting time than GT. Besides, the variance between the customers becomes smaller since we consider the fairness among them. Although IDT has a slight advantage over our A2CMF in minimizing the waiting time, it only focuses on the customer benefit without considering the potential revenue loss and unfair experience among merchants and couriers. As can be seen, A2CMF achieves the best overall performance by improving R_p by 21.3% and reducing ($Gini_c$, G_m) by (9.7%, 6.2%) compared to the second-best baselines.

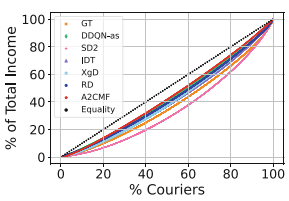

Fig. 8. Rider income lorenz

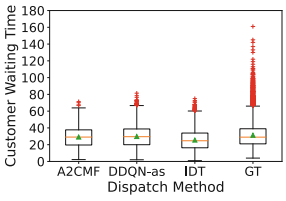

Fig. 9. User waiting time

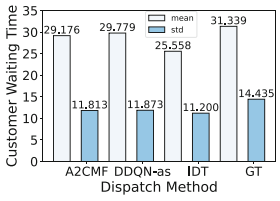

Fig. 10. M_w, D_w comparison

6 Related Works

6.1 Fairness in the Matching Mechanisms

Recently, instant delivery service plays an important role in online ordering and the potential unfairness problem comes into focus. Based on this scenario, researchers seek for matching mechanisms to guarantee fairness. Tom Sühr et al. propose a novel framework to think about not requiring every match to be fair, but rather distributing fairness over time, so they can achieve better overall benefit for all stakeholders [14]. On the other hand, Wang Guang et al. consider fairness as an optimization objective by improving overall efficiency and fairness [16]. And they once leverage greedy algorithm with Pareto improvement to solve multi-objective optimization [15].

6.2 Order Dispatch Mechanisms Based on Reinforcement Learning

Reinforcement learning is widely applied for sequential decision problems and particularly has been adopted for order dispatching in recent years. Ding, Yi et al. build a reinforcement learning model to learn the optimal order dispatching strategies, together with a profit model as the reward function [4]. Considering that instant delivery imposes a strict time deadline, Guo Baoshen et al. propose a Time-Constrained Actor-Critic Reinforcement learning based concurrent dispatch system to enhance long-term overall revenue and reduce overdue rate [5].

7 Conclusion

In this paper, we propose the first multi-sided fairness-aware dispatch system called A2CMF to improve the overall platform revenue and benefit fairness of all stakeholders. We first conduct a data-driven order dispatch pattern analysis, which shows the unfairness of dispatching problem and provides us insights into different notions of fairness among stakeholders. We then formulate the order dispatch as a Markov decision process and use the Advantage Actor-Critic (A2C) algorithm to tackle this problem. The performance of A2CMF is evaluated through a real-world dataset obtained from Eleme including over 1.15 million orders. Experimental results show that our fairness-aware A2CMF effectively

increases the total platform revenue, improves customer service experience, and reduces the benefit gap between couriers and merchants by 9.7% and 6.2%.

Acknowledgements. The authors would like to thank the anonymous reviewers for their constructive and helpful feedback. This work was supported in part by Science and Technology Innovation 2030 - Major Project 2021ZD0114202.

References

1. Abdollahpouri, H., Burke, R.: Multi-stakeholder recommendation and its connection to multi-sided fairness. CoRR abs/1907.13158 (2019)
2. Biega, A.J., Gummadi, K.P., Weikum, G.: Equity of attention: amortizing individual fairness in rankings. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, pp. 405–414. Association for Computing Machinery, New York (2018)
3. Chakraborty, A., Patro, G.K., Ganguly, N., Gummadi, K.P., Loiseau, P.: Equality of voice: towards fair representation in crowdsourced top-k recommendations, FAT* 2019, pp. 129–138. Association for Computing Machinery, New York (2019)
4. Ding, Y., et al.: A city-wide crowdsourcing delivery system with reinforcement learning. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **5**(3), 1–22 (2021)
5. Guo, B., et al.: Concurrent order dispatch for instant delivery with time-constrained actor-critic reinforcement learning. In: 2021 IEEE Real-Time Systems Symposium (RTSS), pp. 176–187 (2021)
6. Lei, H., Zhao, Y., Cai, L.: Multi-objective optimization for guaranteed delivery in video service platform. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2020, pp. 3017–3025 (2020)
7. Li, M., et al.: Efficient ridesharing order dispatching with mean field multi-agent reinforcement learning. In: The World Wide Web Conference, WWW 2019, pp. 983–994 (2019)
8. Li, Y., Zheng, Y., Yang, Q.: Efficient and effective express via contextual cooperative reinforcement learning. In: KDD 2019 (2019)
9. Li, Y., Chen, H., Fu, Z., Ge, Y., Zhang, Y.: User-oriented fairness in recommendation. In: Proceedings of the Web Conference 2021, WWW 2021, pp. 624–632 (2021)
10. Lin, K., Zhao, R., Xu, Z., Zhou, J.: Efficient large-scale fleet management via multi-agent deep reinforcement learning. In: KDD 2018 (2018)
11. McCann, J., Chatley, R.: Fleet management in on-demand transportation networks: using a greedy approach (2018)
12. Patro, G.K., Biswas, A., Ganguly, N., Gummadi, K.P., Chakraborty, A.: Fairrec: two-sided fairness for personalized recommendations in two-sided platforms. In: Proceedings of the Web Conference 2020, WWW 2020, pp. 1194–1204 (2020)
13. Singh, A., Joachims, T.: Fairness of exposure in rankings. In: KDD 2018 (2018)
14. Sühr, T., Biega, A.J., Zehlike, M., Gummadi, K.P., Chakraborty, A.: Two-sided fairness for repeated matchings in two-sided markets: a case study of a ride-hailing platform. In: The 25th ACM SIGKDD International Conference, KDD 2019 (2019)
15. Wang, G., Zhang, Y., Fang, Z., Wang, S., Zhang, F., Zhang, D.: Faircharge: a data-driven fairness-aware charging recommendation system for large-scale electric taxi fleets. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **4**(1), 1–25 (2020)

16. Wang, G., Zhong, S., Wang, S., Miao, F., Dong, Z., Zhang, D.: Data-driven fairness-aware vehicle displacement for large-scale electric taxi fleets. In: ICDE 2021 (2021)
17. Wang, Z., Qin, Z., Tang, X., Ye, J., Zhu, H.: Deep reinforcement learning with knowledge transfer for online rides order dispatching. In: ICDM 2018 (2018)
18. Xiao, L., Min, Z., Yongfeng, Z., Zhaoquan, G., Yiqun, L., Shaoping, M.: Fairness-aware group recommendation with pareto-efficiency. In: RecSys 2017 (2017)
19. Xu, Z., et al.: Large-scale order dispatch in on-demand ride-hailing platforms: a learning and planning approach. In: KDD 2018 (2018)
20. Zhang, L., et al.: A taxi order dispatch model based on combinatorial optimization. In: The 23rd ACM SIGKDD International Conference, KDD 2017, pp. 2151–2159 (2017)
21. Zhang, Y., et al.: Route prediction for instant delivery. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **3**(3), 1–25 (2019)
22. Zhou, M., et al.: Multi-agent reinforcement learning for order-dispatching via order-vehicle distribution matching. In: CIKM 2019 (2019)