Human Aspects in Collaborative Order Picking – What if Robots Learned How to Give Humans a Break?

Niu, Yaxu; Schulte, Frederik

DOI
10.1007/978-3-030-85906-0_59

Publication date
2021

Document Version
Final published version

Published in
Proceedings IFIP International Conference on Advances in Production Management Systems

Human Aspects in Collaborative Order Picking – What if Robots Learned How to Give Humans a Break?

Yaxu Niu[1,2] and Frederik Schulte[2](✉)

[1] Beijing University of Chemical Technology, Beijing 100029, China
2017400145@mail.buct.edu.cn
[2] Delft University of Technology, 2628 Delft, CD, The Netherlands
F.Schulte@tudelft.nl

Abstract. Human aspects in collaboration of humans and robots, as common in warehousing, are considered increasingly important objectives in operations management. In this work, we let robots learn about human stress levels based on sensor data in collaborative order picking of robotic mobile fulfillment systems. To this end, we develop a multi-agent reinforcement (MARL) approach that considers human stress levels and recovery behavior next to traditional performance objectives in the reward function of robotic agents. We assume a human-oriented assignment problem in which the robotic agents assign orders and short breaks to human workers based on their stress/recovery states. We find that the proposed MARL policy reduces the human stress time by up 50% in comparison to the applied benchmark policies and maintains system efficiency at a comparable level. While the results may need to be confirmed in different settings considering different types of humans aspects and efficiency objectives, they also show a practicable pathway to control stress levels and recovery for related problems of human-robot collaboration, inside and outside of warehousing.

Keywords: Order picking · Robotic mobile fulfillment systems · Human aspects · Multi-agent reinforcement learning · Human-robot collaboration · Sensor data · Recovery

1 Introduction

Human-oriented collaboration between humans and robots is widely considered one of the greatest challenges in the final steps of the 4th Industrial Revolution and an anticipated central question of the 5th Industrial Revolution. Order-picking in robotic mobile fulfillment systems (RMFS) is one of the applications in which human-robot collaboration is already a pivotal element of today's working reality. Various authors have recognized and addressed the issue with a respective

Thanks for the support of China Scholarship Council (CSC).

Published by Springer Nature Switzerland AG 2021
A. Dolgui et al. (Eds.): APMS 2021, IFIP AICT 632, pp. 541–550, 2021.
https://doi.org/10.1007/978-3-030-85906-0_59

survey paper [1], a conceptual framework for the integration of humans aspects in planning approaches of ordering picking [2], or specific operational models (e.g., [3]). Niu et al. [4] have proposed a decision support approach that enables robots to learn a human-oriented assignment policy in collaborative order picking based on a discomfort model introduced by Larco et al. [5]. On top of that, Sedighi Maman et al. [6] have recently demonstrated how sensor data from wearable devices can be used to detect the stress levels and recovery of workers. Nonetheless, to leverage these findings for human well-being in collaborative operations, the two methods would need to be integrated in an approach that learns to take (assignment) decisions based on sensor data indicating human stress and recovery levels.

In RMFSs, most related works focus on different variants of operational assignment problems considering a general assignment between work stations and robots [7], shelves to storage assignment [8], order and storage assignment [3], velocity-based storage assignment [9], and robots to shelves assignment [10]. Apart from that, workstation location problems [11], pod travel times under different lane characteristics [12], fleet sizing [13], path planning [7], and order batching and shelf sequencing [14] are considered. The most common methods are statistical models, analytical and queuing models as well as optimization models. Only [13] propose a MARL negotiation scheme and apply it to an order picking example application, and [11] develop a discrete event simulation framework for RMFS. Among these references, only [7] consider human aspects using the proxy of the workers' handling speed.

In this work, we propose a multi-agent reinforcement learning (MARL) approach in which robotic agents effectively learn to consider human needs for recovery based on wearable sensor data, next to established objectives such as minimum processing times. This assumed human-oriented assignment with its sensor-based decisions is illustrated in Fig. 1 based on the layout of the underlying RMFS. The proposed policy considers short breaks for workers that are implemented via assignment decisions. For the conducted experimental study, this paper develops four different policies that are commonly deployed in order picking. The proposed MARL reduces the total stress time for humans by up 50 % during the collaborative order picking process without significantly compromising the system efficiency. While it appears probable that the evaluated policy can be further improved in terms of human aspects as well as operational efficiency, the experiments confirm that the MARL approach enables robotic learning with respect to human aspects and (multiple) other objectives. Since the general characteristics of the considered human-oriented assignment problem resemble many other operational problems including human-robot collaboration, the MARL approach also be adopted in different and new problem settings.

Subsequently, we present the human-oriented assignment problem in order picking of an RMFS (in Sect. 2), the proposed MARL approach (in Sect. 3), the experimental study with results (in Sect. 4), and a conclusion with open issues for future work (in Sect. 5).

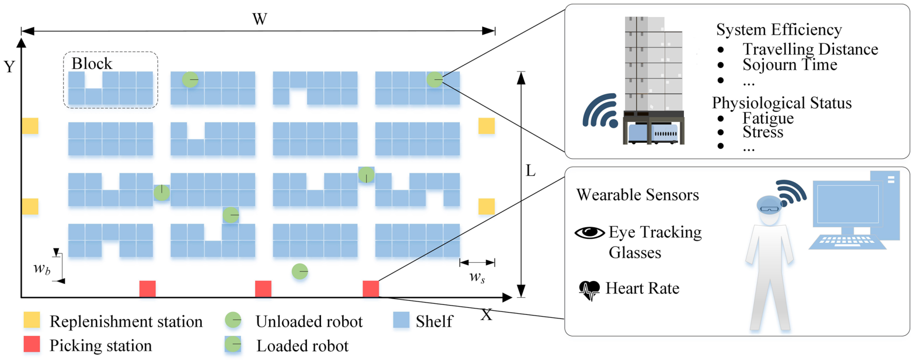

Fig. 1. Layout of the RMFS illustrating robotic assignment decision points in which human sensor data is considered

2 Problem Description

In this section, the underlying RMFS in this paper is described and the assumptions used are presented in Sect. 2.1. Then, in Sect. 2.2, the human-oriented robot assignment problem is defined, both considering system efficiency and human well-being objectives.

2.1 The Robotic Mobile Fulfillment System

The layout of RMFS is shown in Fig. 1, where shelves are organized as rectangular blocks in the storage area. The workstations are situated along the boundary of the warehouse. The available robot moves underneath the shelves from the dwell location toward the targeted shelf. Then robot lifts the movable shelf and transports it to the designated workstations along the aisles and cross-aisles, and queues for its turn if the worker is busy. Finally, the order is fulfilled and the robot transports the completed shelf to its previous storage position.

In order picking activities, pickers are engaged in intensive and repetitive manual handling activities, which will inevitably lead to the accumulation of physiological fatigue and the generation of stress, which will affect the efficiency of the system and may even endanger health. In RMFSs, frequent and repetitive item picking, handling, and packing activities will contribute to the accumulation of fatigue, which will result in a decrease in efficiency and an increase in reaction time. Besides, it has been demonstrated that heart rate monitoring has turned out to be an efficient way to realize fatigue detection. As presented in [6], fatigue detection is mainly comprised of the following steps: (1) selected sensors' data are preprocessed for feature generation; (2) then processed data are used to train the statistical and data analytic model for distinguishing between fatigue and non-fatigued state; (3) the trained models are evaluated based on accuracy, sensitivity, etc.; (4) finally, the best-trained model is evaluated.

The stress of the order pickers mainly results from the increase in workload or the occurrence of unexpected events. This article assumes that the generation of stress is caused by an excessive workload. Moreover, the relationship between stress and performance follows the Yerkes-Dodson law, represented as an inverted-U shape where high performance can only be guaranteed within a moderate stress level. It's interesting that if stress occurs, the PD tends to increase, on the contrary, the PD remains small [15]. Therefore, stress detection, in this paper, referred to an on-line stress detection approach proposed by [15], which processes pupil diameter derived features

This paper mainly focuses on seeking an intelligent robot assignment policy to cope with dynamic changes in pickers' efficiency due to fatigue, stress, and short-break recovery, balancing system efficiency, and human well-being objectives. The studied assignment problem assumes that the previous decision problems, such as order assignment and picking shelf selection, have been solved. Thus, the next target shelf has already been determined. For the model formulation, some assumptions, which are reasonable in real RMFS, are first listed as follows.

– It is assumed that all the robots are busy and there are always retrieval orders waiting in the external queue.
– Each robot in the system executes its tasks independently and only wirelessly communicates with workstations to obtain environmental information.
– Robots are scheduled based on the First-Come-First-Served (FCFS) policy.
– The orders are single-line orders, which are common and account for a large number of E-commerce orders.

2.2 Human-Oriented Assignment Problem of the Robot to Workstations

In a typical RMFS, there always exist numerous unfulfilled order tasks. Suppose that robots are dealing with $\mathbb{O}$ fixed sequence orders scattered among different shelves parking in different positions in the storage area, and then will be assigned to pickers with different handling speed T_{w_k} to fulfill the retrieval process. This robot assignment problem is modeled as a Markov Game model, in which all the robots make assignment decisions following a joint policy π, $\pi \in \mathit{\Pi}$ to service a sequence of order tasks.

The robot r_j at the dwell position receives an order task located at the position p_{o_i} and can realize retrieval transaction by assigned to one of the workstations. Let $\mathbb{A} = [0, 1, ..., n_{w_l}]$ be the set of all the assignment decisions. When n_{w_k} is selected in the set $\mathbb{A}^0 = [1, ..., n_{w_l}]$, it means the robot r_i is assigned to the corresponding workstation w_l, whereas 0 means that the robot is working on current order task not making assignment decisions at the moment. The available robot will make assignment decisions from two aspects: system efficiency and picker condition. On the one hand, the robot evaluates the expected time to reach all workstations based on the target shelf position, to measure the efficiency impact of selecting different workstations. On the other hand, the fatigue and stress level of each picker will be transmitted to the current robot via the wireless device after processing the physiological signal collected by the wearable device at the workstation. The picker status will be an important factor in robot decision-making to achieve a human-oriented assignment policy. Based on the picker's physiological signals, the robot will avoid selecting fatigued/stressed pickers to fulfilled the order task. At the same time, the current workload of each workstation, such as the length of the waiting queue and the number of shelves to the current workstation, is also an important basis for the assignment decision. It not only affects the sojourn time of the shelf at the workstation and therefore contributes to the efficiency of the system, but also affects the physiological state of the picker (over high and low workloads will lead to stressed-out and passive on picker respectively). Based on the information obtained above, the robot makes assignment decisions for balancing system efficiency and worker fatigue and stress. The assignment decision is based on the system state information that includes the robot state vector, order information vector, and human information vector. Figure 2 shows how robots and pickers cooperate to complete shared tasks.

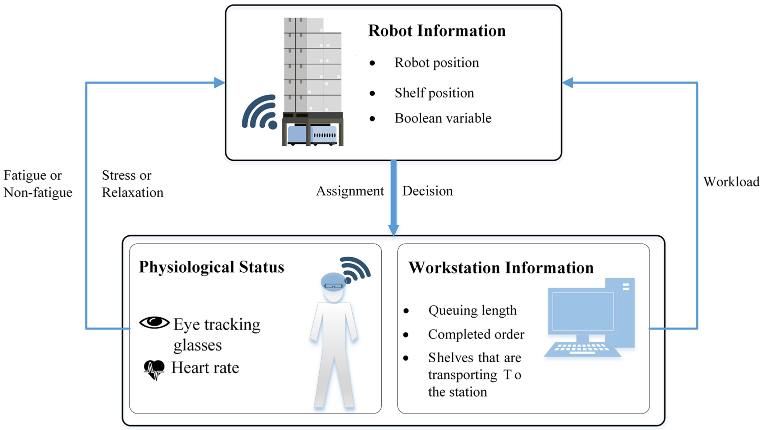

Fig. 2. The mechanism of robots and pickers cooperation for sharing tasks

Specifically, in the setting of this article, the robot can determine its current position p_{r_j} by scanning the QR code on the warehouse ground. Considering the current time step, the position can be expressed as $p^t_{r_j}$. For each robot, a boolean b is used to indicate task status. If the robot is working on the current order task, $b = 0$, otherwise $b = 1$. Each workstation information vector is defined as $WI_k = [N^k_{que}, N^k_{tran}, N^k_{comp}]$, where N^k_{tran} represents the number of orders completed within the time interval T_{int} between two assignment decisions time step recorded by the workstation. The workstation also records the current queue length N^k_{que} and the number of shelves being transported N^k_{tran} to the workstation. As shown in Fig. 1, based on the pickers' real-time physiological signals collected from wearable sensors, signal processing, and data analysis methods allow us to detect the current picker's stress state S_{stress} and fatigue state $S_{fatigue}$. Thus, each human information vector can be defined as $HI_k = [S^k_{stress}, S^k_{fatigue}]$. With all these required vectors, the system state at the time t is defined as $S_t = [RS_t, WI_t, HI_t]$, where $WI_t = [WI^t_0, ..., WI^t_l]$ and, $HI_t = [HI^t_0, ..., HI^t_l]$.

Based on the current system status, all robots follow the joint assignment policy, which will contribute to the total fulfilled orders and the stress level of the human picker who is assigned to complete the order task. Meanwhile, robots will receive the corresponding penalties on time cost and discomfort. Assuming time cost and discomfort penalties are linear, the penalty function at state S_t with joint action is given by

$$R_t(S_t, n_t) = -T_{int} - \beta T_{stress} \tag{1}$$

where β is the weight of the element to evaluate the importance of this item. This paper aims to determine the joint policy π^* for all robots which respond to the efficiency changes of pickers due to fatigue, stress, and short breaks, thereby contributing to system efficiency and worker well-being, namely realizing equilibrium, and maximizes the expected cumulative penalty.

$$J(\pi^*) = max_{\pi \in \Pi}(\sum_{o_i=0}^{m} R_t(S_t, n_t)) \tag{2}$$

3 Multi-agent Reinforcement Learning Approach

In this section, a robot-based multiagent reinforcement learning (MARL) method for the robot assignment problem that aims at obtaining a human-oriented assignment policy considering both system efficiency and the pickers' discomfort is presented. In the RMFS, learning in a multi-robot environment is inherently complex. In this paper, we propose a Value-Decomposition Network (VDN) approach to solve the assignment problem, which is a centralized training and decentralized execution framework, allowing the policy to use other agents' local information for training, and execute each robot's action only via individual observations [16]. The VDN value function for the system is estimated under the assumption that the value decomposition networks learn to decompose the total Q value function into value functions across agents that condition only on robot individual observations, namely

$$Q_{total}(S_t, n_t; \boldsymbol{\theta}) = \sum_{j=0}^{n} Q_{r_j}(S_t^{r_j}, n_{w_k}; \theta^{r_j}) \tag{3}$$

where θ are the parameters of the neural network, which is introduced to deal with the "curse of dimensionality" caused by enumerating all the combinations of robot location, picker information, and workstation information. Finally, following the independent deep Q-learning rules, the problem is solved by minimizing the following loss function for VDN.

$$L(\boldsymbol{\theta}) = \sum_{\omega=1}^{M}\sum_{j=0}^{n}(R_\omega + \gamma max_{n'_{w_k}\in\mathbb{A}} Q(S_\omega^{'r_j}, n'_{w_k}, \theta^-) - Q(S_\omega^{r_j}, n_{w_k}, \theta))^2 \tag{4}$$

4 Experimental Results

In this section, the performance of the proposed human-oriented MARL assignment policy (HOAP) is compared with the other three traditional robot assignment policies, namely random assignment policy (RAP), nearest assignment policy (NAP), and shortest-queue assignment policy (SAP). The simulation experiments are performed on a warehouse equipped with 2 picking stations, investigating the system efficiency and human discomfort during 30 min work. Beside, if the size of the robot fleet exceeded the operator's capability, the operator will always face stress, contributing to a total stress time increase, while the size of the robot fleet is too small, unable to exert the operator's performance, resulting in a decrease in system efficiency. Therefore, in this article, we set 16 robots in the experiment which is a good equilibrium point(number of robots necessary compared to the number of operators). An aisle and two cross-aisles are deployed in the storage zone, dividing the zone into six 5×2 rectangle blocks. The handling speed of workstations w_0 and w_1 is set to 6 shelves/min and 10 shelves/min, respectively, to consider the different speeds for different pickers. Besides, at the beginning of the experiment, we assume that the picker working at station w_1 has been working for a while and will have a short break of 10 min after 5 min to get recovery.

During the experiment, the agents learn for 500 episodes with the discount factor is set to 0.95, and the exploration rate decreases from 1 to 0.01 in 6000 time steps. And the number of nodes of the one layer in the Q network is 128. Based on the learned policy, a comparison among all assignment policies in terms of total completed orders and stress time for 100 run when $t = 20$ min and $t = 200$ min are presented in Fig. 3 and Fig. 4.

It can be seen from the figures that the proposed human-oriented MARL policy and the shortest assignment policy always outperform the other two assignment policies, and have little difference in system efficiency, which both completed nearly 330 and 290 orders when $t = 20$ min and $t = 200$ min. Comparing Fig. 3 and Fig. 4, we can see that at $t = 200$ min, the efficiency of all policies is lower than that of their own performance at $t = 20$ min. This is because pickers can maintain a high working efficiency at $t = 20$ min, fatigue and stress do not have a significant impact on efficiency, while picker fatigue accumulates with the increase of working time and leads to a decrease in efficiency. Although in the experiment one of the pickers, w_1, arranged a short break to relieve fatigue and work stress, not enough to fully recover. In terms of the average total stress time, the proposed learning-based method has a great improvement in reducing pickers' stress time, whereas random assignment, nearest assignment, and shortest assignment have a higher total stress time. The total stress time of the learning-based policy is only 50% to 65% of the stress time of the other three policies. This phenomenon can be explained by the fact that these three policies only mechanically assign orders to the workstation based on rules, without considering the status of the pickers. However, the proposed policy, while considering the system efficiency, is also aware of the real-time status of pickers and makes assignment decisions to reduce the duration of total stress time. The MARL can greatly alleviate the stress of the order pickers without significantly sacrificing system efficiency. Therefore, compared with a traditional rule-based allocation method, the method proposed in this article not only considers the total amount of completed orders, but also comprehensively considers the fatigue and stress levels of the order pickers, achieving the bi-objective solution for the system efficiency and worker well-being.

Fig. 3. Average completed orders and total stress time when $t = 20$ min

Fig. 4. Average completed orders and total stress time when $t = 200$ min

5 Conclusion

The consideration of human aspects in the collaboration between humans and robots has been widely recognized as one of the major challenges in the era of the 4th Industrial Revolution. Also, human aspects of order picking in warehousing have received significant attention in recent research. However, human well-being has hardly been considered in collaborative order picking (of humans and robots), as common in robotic mobile fulfillment systems. This work has proposed a multi-agent reinforcement learning approach that considers wearable sensor data to incorporate human stress levels and recovery behavior for a human-oriented assignment problem in robotic mobile fulfillment systems. The results show that the developed approach reduces human stress times by nearly 50 % and achieves a similar system efficiency as three considered benchmark policies. In this way, our research extends existing work on wearable sensors to detect human stress [6] and human aspects in collaborative order picking [4] by introducing a new method for robotic learning in RMFS and by considering short breaks to control human stress levels and recovery in order picking of RMFS. Learning this, robots grow into the role of caring colleagues for human co-workers who, for instance, tell humans when they need a break. Moreover, the underlying human-oriented assignment problem resembles a growing amount of related problems, and the proposed MARL approach may therefore also be adopted in those other domains. Nonetheless, there are a few extensions to be considered for a more coherent understanding of this new form of human-robot collaboration. In this work, two pickers are considered but an extension to more pickers will probably create more options to balance the workload. Similarly, the robots maybe also pause when all workers show elevated stress levels. Future work will, furthermore, define different options to connect the theory of sensor-based stress and recovery curves to collaborative operations research models. Finally, we also aim to explore further use cases in problems of multi human-robot collaboration, inside and outside of warehousing.

References

1. Grosse, E.H., Glock, C.H., Neumann, W.P.: Human factors in order picking: a content analysis of the literature. Int. J. Prod. Res. **55**(5), 1260–1276 (2017)
2. Grosse, E.H., Glock, C.H., Jaber, M.Y., Neumann, W.P.: Incorporating human factors in order picking planning models: framework and research opportunities. Int. J. Prod. Res. **53**(3), 695–717 (2015)
3. Merschformann, M., Lamballais, T., De Koster, M., Suhl, L.: Decision rules for robotic mobile fulfillment systems. Oper. Res. Perspect. **6**, 100128 (2019)
4. Niu, Y., Schulte, F., Negenborn, R.R.: Human aspects in collaborative order picking-letting robotic agents learn about human discomfort. Procedia Comput. Sci. **180**, 877–886 (2021)
5. Larco, J.A., De Koster, R., Roodbergen, K.J., Dul, J.: Managing warehouse efficiency and worker discomfort through enhanced storage assignment decisions. Int. J. Prod. Res. **55**(21), 6407–6422 (2017)

6. Maman, Z.S., Chen, Y.J., Baghdadi, A., Lombardo, S., Cavuoto, L.A., Megahed, F.M.: A data analytic framework for physical fatigue management using wearable sensors. Expert Syst. Appl. **155**, 113405 (2020)
7. Zou, B., Gong, Y., Xu, X., Yuan, Z.: Assignment rules in robotic mobile fulfilment systems for online retailers. Int. J. Prod. Res. **55**(20), 6175–6192 (2017)
8. Weidinger, F., Boysen, N., Briskorn, D.: Storage assignment with rack-moving mobile robots in kiva warehouses. Transp. Sci. **52**(6), 1479–1495 (2018)
9. Yuan, R., Graves, S.C., Cezik, T.: Velocity-based storage assignment in semi-automated storage systems. Prod. Oper. Manage. **28**(2), 354–373 (2019)
10. Roy, D., Nigam, S., de Koster, R., Adan, I., Resing, J.: Robot-storage zone assignment strategies in mobile fulfillment systems. Transp. Res. Part E Logistics Transp. Rev. **122**, 119–142 (2019)
11. Lamballais, T., Roy, D., De Koster, M.: Estimating performance in a robotic mobile fulfillment system. Eur. J. Oper. Res. **256**(3), 976–990 (2017)
12. Wang, K., Yang, Y., Li, R.: Travel time models for the rack-moving mobile robot system. Int. J. Prod. Res. **58**(14), 4367–4385 (2020)
13. Zhou, L., Yang, P., Chen, C., Gao, Y.: Multiagent reinforcement learning with sparse interactions by negotiation and knowledge transfer. IEEE Trans. Cybern. **47**(5), 1238–1250 (2016)
14. Boysen, N., Briskorn, D., Emde, S.: Parts-to-picker based order processing in a rack-moving mobile robots environment. Eur. J. Oper. Res. **262**(2), 550–562 (2017)
15. Ren, P., Barreto, A., Huang, J., Gao, Y., Ortega, F.R., Adjouadi, M.: Off-line and on-line stress detection through processing of the pupil diameter signal. Ann. Biomed. Eng. **42**(1), 162–176 (2014)
16. Sunehag, P., et al.: Value-decomposition networks for cooperative multi-agent learning. arXiv preprint arXiv:1706.05296 (2017)