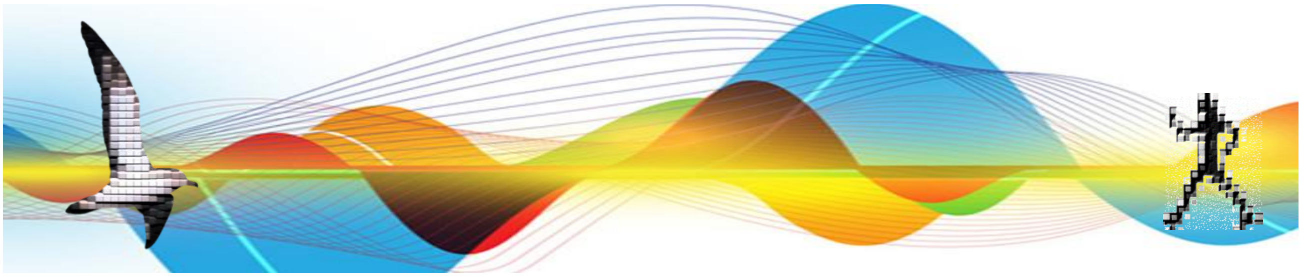
# Online Proceedings of BiRD 2023



# 2nd International Workshop on

## Behavior analysis and Recognition for knowledge Discovery

# BiRD 2023 Special Talk: Learning Interaction Rules from Multi-Agent Trajectories

Keisuke Fujii
*Nagoya University*
Nagoya, Aichi, Japan
fujii@i.nagoya-u.ac.jp

***Abstract*—Understanding real-world multi-agent motions is a fundamental issue in engineering, biology, and human behavioral science. However, in real-world multi-agent systems, the rules behind their complex movements are often unknown. In such cases, it would be effective to estimate the unknown parts using machine learning from measured data. In this talk, I will first introduce a machine learning method for estimating interaction rules from movement trajectories in multiple species of animals and a theoretical model of animal behaviors. We adopt an approach for augmenting incomplete multi-agent behavioral models described by time-varying dynamical systems with neural networks. For efficient and interpretable learning, our model leverages theory-based architectures separating navigation and motion processes, and the theory-guided regularization for reliable behavioral modeling. Next, we introduce a causal inference method for estimating time-varying individual treatment effects (ITEs) from limited data for multi-agent trajectories in different domains such as biological movement simulators, autonomous driving simulators, and team sports. Our model leverages graph variational recurrent neural networks and theory-based computation with domain knowledge for the ITE estimation framework based on long-term prediction of multi-agent covariates and outcomes. These frameworks would provide insights into learning approaches of multi-agent rules from real-world trajectory data.**

# Prediction of Bat Flight Path During Obstacle Avoidance by Imitation Learning


Shoko GENDA
*Graduate School of Life and Medical Science*
*Doshisha University*
Kyotanabe, JPN
ctuh1006@mail4.doshisha.ac.jp

Yu TESHIMA
*Graduate School of Life and Medical Science*
*Doshisha University*
Kyotanabe, JPN
yuteshima18@gmail.com

Taku KAWAMURA
*Graduate School of Life and Medical Science*
*Doshisha University*
Kyotanabe, JPN
ctug1013@mail4.doshisha.ac.jp

Daichi OHARA
*Faculty of Life and Medical Science*
*Doshisha University*
Kyotanabe, JPN
cgsd1075@mail4.doshisha.ac.jp

Keisuke FUJII
*Graduate School of Informatics*
*Nagoya University*
Nagoya, JPN
fujii@i.nagoya-u.ac.jp

Shizuko HIRYU
*Graduate School of Life and Medical Science*
*Doshisha University*
Kyotanabe, JPN
shiryu@mail.doshisha.ac.jp



***Abstract*— Bats use echolocation, emitting ultrasound and listening to the echoes to perceive their environment instead of visual information. As the space perceived by bats is different from the visual space and they have complex and diverse behavioral patterns, it is difficult to develop a bat navigation model that accounts for task and environmental interactions. In this study, we developed a predictive model of bat flight paths during obstacle avoidance by imitation learning. In the behavioral experiments, we recorded the flight trajectories of five eastern bent-winged bats (*Miniopterus fuliginosus*) during obstacle avoidance using a motion-capture system. Two obstacle conditions were provided: one with obstacles randomly placed in a closed space and the other with obstacles regularly placed in a closed space. The predictive model of the bat flight path was created by imitation learning with a recurrent neural network in the randomly placed obstacle condition with the bat's behavior as output. Simulations using the predictive model confirmed that the flight path could be predicted, even in an environment with unknown obstacle placement that had not been learned beforehand.**



***Keywords*— *echolocation, bat, obstacle avoidance, imitation learning, flight path, predictive model***


## I. Introduction

Echolocating bats emit ultrasound pulses and listen to the returning echoes from objects to understand their surrounding environment [1]. Bats are thought to navigate flight paths to avoid obstacles efficiently and attack prey based on the echo space perceived by sound. A previous study focused on the navigation of foraging bats toward their prey, and mathematical modeling of their flight paths suggested that bats select efficient flight paths [2]. However, the navigation of bats is so diverse and complex that it is difficult to derive simple mathematical models that take into account the interaction of each task and environment from behavioral data. Furthermore, the generation of navigation models is made more complex as the space perceived by bats is different from the space that we perceive visually [3]. In this study, we created a predictive model of flight paths during obstacle avoidance by imitation learning based on measured data from behavioral experiments. The proposed model was then evaluated in simulations comparing the predicted flight paths with the measured flight paths of bats.

## II. Methods

### A. Behavioral Experiments

In the behavioral experiments, we recorded the three-dimensional flight trajectories of five eastern bent-winged bats (*Miniopterus fuliginosus*) at 100 frames/s using a motion-capture system with 16 cameras. The bats flew in a flight room measuring 4.5 m (L) × 4.5 m (W) × 2.5 m (H), which was divided on one side by a net. From 12 to 20 obstacles (plastic chains 0.03 m in diameter) were suspended from the ceiling. A total of six obstacle layouts were prepared: three with randomly arranged chain obstacles (Fig. 1A) and three with regularly arranged chain obstacles (Fig. 1B). We changed the starting point from which the bats commenced their flight and measured the flight path of each individual bat for 5 min.

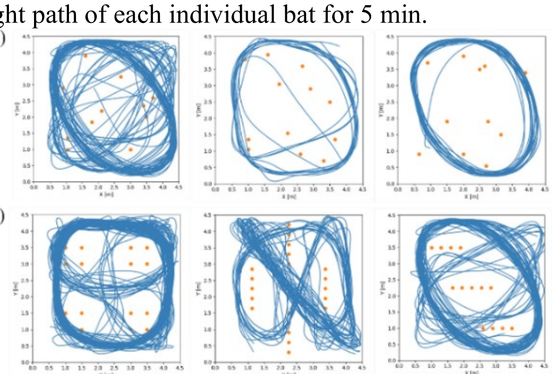



Fig. 1. Top views of the flight trajectory of one bat in six obstacle layouts with chains arranged randomly for training (A) and regularly for test (B).

### B. Data sets

The data set for imitation learning was created from the measured flight trajectories obtained in the random obstacle condition shown in Fig. 1A. The data set consisted of action and state as time series data. The action data consisted of the location, velocity, and turning angle of the bat. For state data, a fan-shaped area within 5.0 m delimited by ±30°, the sound directivity range of *M. fuliginosus* [4] was divided into 201 sensing lines in 0.3° increments (Fig. 2). In addition, the normalized distance from the bat's position to the target (room wall or net, and chain obstacle) on the sensing line was calculated every 0.1 s (one step). One episode was defined as 80 steps, and the training data consisted of 217 episodes of data taken from three random obstacle conditions. The test data consisted of 30 episodes of data from three regular obstacle layouts.

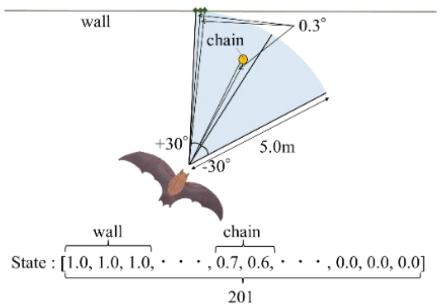



Fig. 2. Example of state data.

### C. Model

We created a predictive model of flight paths by imitation learning on the data set of the regular obstacle condition. We used a recurrent neural network (RNN) developed in a previous study [5]. The input dimension of the RNN was 206, the hidden layer dimension was 64, the hidden state dimension was 100, and the output dimension was 2. The accuracy of the model was defined as the difference between the measured and predicted flight velocities. This is because flight coordinates are a position-dependent parameter, whereas velocity is a position-independent parameter. We used 10% of the training data in the validation trial to check the performance of the temporary model. In this obstacle space, the bats fly 1.5 laps in about 8 s. Therefore, we wanted to predict 1 lap and set the prediction interval to 6 s. The input time was set to 2 s based on previous study [5].

## III. Results

Figure 1 shows the flight trajectories of one bat in the six obstacle conditions. The bats tended to avoid obstacles by repeatedly following the same path, and the pattern of flight paths was similar for all individuals, depending on the layout of the obstacles.

The average loss for all test trials (30 episodes) was 0.78 ± 0.48 m/s, which was comparable to the variation in the flight path of the bats in this study (0.85 m/s), and was therefore considered to be adequate for prediction accuracy (Table I). Furthermore, comparison with a previous study that predicted the in-game movements of soccer players [5] suggested that our method is sufficiently versatile to predict the movements of animals as well as humans (Table I).

Figure 3 shows the predicted flight paths in comparison with measured flight path data in the three regular obstacle conditions. In the test, the model was able to predict the location of bats with an accuracy of 0.42 ± 0.28 m. This may be sufficiently accurate considering the wingspan of the bat is about 0.25 m and the displacement caused by the flapping of wings. The predicted flight path is not necessarily a copy of the input, because the model predicts velocity, and there is variability in the predicted velocity (Table 1). In all unlearned spaces, the predicted flight paths tended to be similar to the measured flight paths. This suggested for the first time that bats do not have space-dependent or individual-dependent flight strategies but, rather, flight strategies common to the species.

TABLE I. Measured Velocity, Predicted Velocity, and Loss Between Them in This and Previous Studies

| | Soccer[5] | Bat |
|---|---|---|
| Measured velocity [m/s] | 2.28 ± 1.57 | 3.05 ± 0.85 |
| Predicted velocity [m/s] | 2.64 ± 1.92 | 2.92 ± 1.07 |
| Losses [m/s] | 1.71 ± 0.23 | 0.78 ± 0.48 |

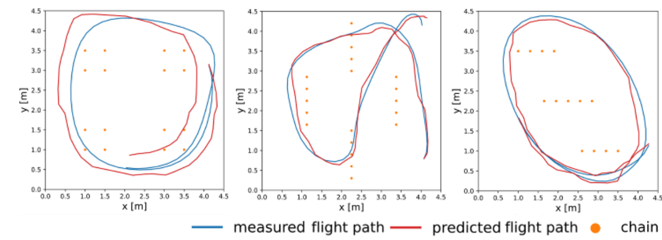



Fig. 3. Top views of measured and predicted flight paths for the test trials.

## IV. Conclusion

In this study, we focused on the obstacle avoidance behavior of bats and developed a predictive model for flight paths. In the behavioral experiments, all bats stabilized on flight trajectories specific to the obstacle layout. Then, flight paths were modeled from the measured data by imitation learning. The model was able to predict flight paths even in environments with unknown obstacle distributions, suggesting that this may be a general-purpose model, independent of the environment. The results are expected to benefit other organisms and robots, because this model can imitate real behavior based on forward data alone, without information about the environment.

In future, it will be necessary to incorporate bat echolocation tactics, such as acquiring information on obstacles according to the direction of the pulse emitted by the bat, i.e., not from the direction of flight, but from the timing of the bat's pulse emission. It will be necessary to represent the mechanism of acquisition of environmental information by echolocation in a model to understand the navigation tactics of bats.

## Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers 21H05295 and 22H01503.

## References

[1] D. R. Griffin, "Listening in the dark: the acoustic orientation of bats and men", Yale University Press, 1958.

[2] E. Fujioka, I. Aihara, M. Sumiya, K. Aihara, S. Hiryu, "Echolocating bats use future-target informatin for optimal foraging", PNAS 113.17, pp. 4848-4852, 2016.

[3] S. Greif, B. M. Siemers, "Acoustic mirrors as sensory traps for bats", Science 357, pp. 1045-1047, 2017.

[4] L. Jakosen, J. M. Ratcliffe, A. Surlykke, "Convergent acoustic field of view in echolocating bats", Nature 493, pp. 93-96, 2013.

[5] K. Fujii, N. Takeishi, Y. Kawahara, K. Takeda, "Policy Learning with Partial Observation and Mechanical Constraints for Multi-person Modeling", arXiv preprint arXiv:2007.03155, 2020.

# Work Recognition Based on Product States, Processing Parts, and Skeleton Data

Hirotomo Oshima
Corporate Manufacturing engineering Center
Toshiba Corporation
Yokohama, Japan
hirotomo1.oshima@toshiba.co.jp

Takanori Yoshii
Corporate Manufacturing engineering Center
Toshiba Corporation
Yokohama, Japan
takanori.yoshii@toshiba.co.jp

Masamitsu Fukuda
Corporate Manufacturing engineering Center
Toshiba Corporation
Yokohama, Japan
masamitsu1.fukuda@toshiba.co.jp

Takuya Maekawa
Graduate School of Information Science and Technology
Osaka University
Osaka, Japan
maekawa@ist.osaka-u.ac.jp

Yasuo Namioka
Corporate Manufacturing engineering Center
Toshiba Corporation
Yokohama, Japan
yasuo.namioka@toshiba.co.jp

***Abstract*— To improve manufacturing productivity, manufacturing processes which are done by workers or machines are continuously monitored, and some causes of work delay against the plan are traced and analyzed. However, an assembly process consists of ten or more elemental works, and the data of the work are recorded manually. That is an issue in terms of work efficiency and personnel resources. We thus propose a method which recognizes elemental works of an assembly process by using Graph Neural Networks. This method enables to estimate by fusing information about product states, processing parts of worker's hands, and worker's skeleton data of each elemental work. Accuracy and validity of this method was investigated with actual data which are collected at an assembly process.**

**Keywords— Work Recognition, Assembly process, Product states, Processing parts, Skeleton data, GNN**

## I. Introduction

With the development of AI and IoT technologies, manufacturing sites are increasingly becoming smart factories. Especially, understanding work progress is important because it is directly related to the improvement of productivity.

It is necessary to recognize what workers do accurately for understanding work progress. Today, there is a lot of researches on action recognition technologies based on the information of worker's movements[1][2]. However, it is difficult to identify detailed works (hereafter referred to as elemental works) accurately using only the information of worker's movements.

Therefore, in this paper, we aim to recognize elemental works by fusing information about worker's movement, product states, worker's processing parts data. The contribution of this study is to develop a unified graph neural network-based method that efficiently fuses the three information.

## II. Related Research

Various methods for work recognition have been proposed. For example, methods that utilize worker's data acquired from wearable devices have been reported[1]. Recently, because a method that can extract human skeleton data accurately and automatically has been established, research on motion recognition using GNN(Graph Neural Network) has attracted attention[2].

## III. Proposed Method

In this paper, we propose a GNN-based method that recognizes elemental works of an assembly process using information of product states, worker's processing parts, and worker's skeleton data integratedly. When humans identify elemental work, they make integrated judgements based on product states, worker's processing parts, worker's posture and position. Therefore, we considered applying GNN to fuse three types of information: product states, worker's processing parts, in addition to worker's skeleton data.

Figure1 shows an overview of the proposed method. From the captured work video, the method extracts worker's skeleton data, product states, and worker's two-hand processing parts in parallel, and estimates elemental work by inputting the extracted values into each predefined graph structure. Skeleton graph structure was defined based on prior studies. The product state can be represented as a graph of state transitions because the appearance of the product changes over time, and the worker's two-hand processing parts can also be represented as a graph by defining in advance where the worker will work.

## IV. Evaluation

### A. *Target Assembly Process/ Evaluation Method*

We targeted an assembly process of an air-conditioner which consists of sixteen different elemental works, and conducted the evaluation experiments which estimate elemental works from recordings of the work being performed by five different workers.

The following two methods were used to estimate the elemental works in the target assembly process. We analyzed the

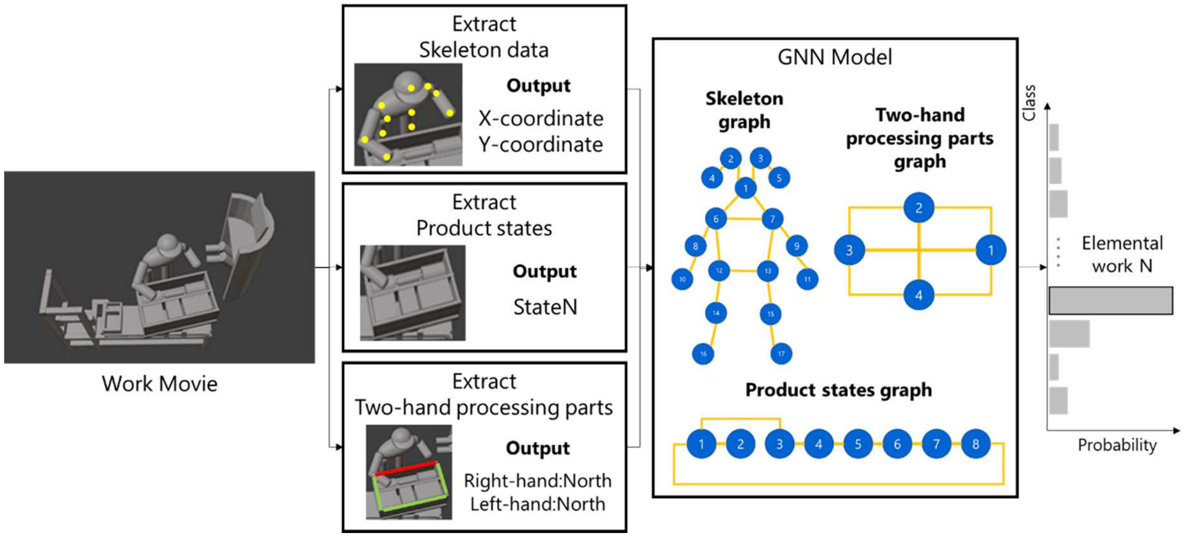



**Figure 1 Overview of the Proposed Method**

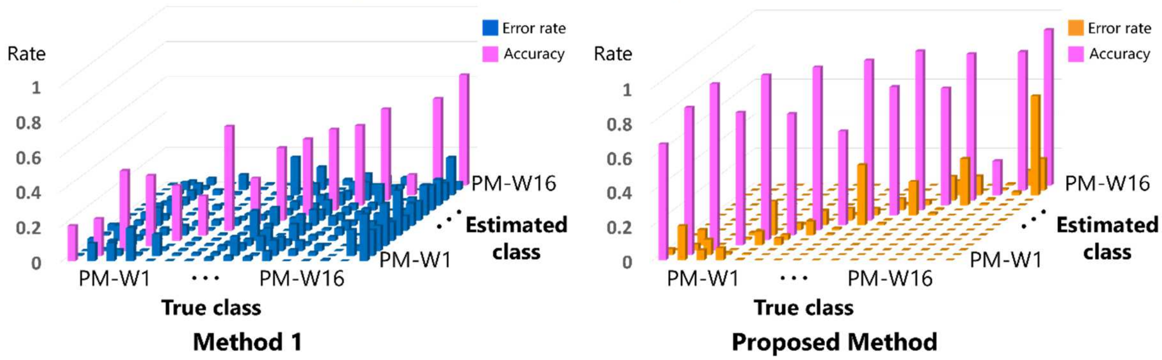



**Figure 2 Confusion matrix**

effect of adding the product states and processing parts on the estimation results.

Method1. Estimation of elemental work by applying only skeleton data to GNN

Method2. Estimating the elemental works by applying the product states, the worker's two-hand processing parts, and the worker's skeleton data to GNN (Proposed Method)

### *B. Results*

Comparing Method 1 to the Proposed Method, the average F-measure was 0.3951 for Method 1, 0.7833 for the Proposed Method, and the Proposed Method was significantly better than Method 1. And, different trends were observed in each estimation result. To explain the difference in estimation accuracy between Method 1 and the Proposed Method, the distribution of estimation results is shown in Figure 2. It can be seen that Method 1 incorrectly estimates various other elemental works. This is considered to be due to the fact that Method 1 cannot distinguish elemental works from other elemental works done in similar postures and positions, because it estimates elemental works only from the worker's skeleton data. On the other hand, the Proposed Method is generally able to correctly estimate the elemental works, and it can be confirmed that the misestimation is concentrated on the adjacent elemental work. This indicates that, even if estimating other elemental works done in similar postures and positions, it is possible to narrow down choices of elemental works by integratedly estimating using data of the product states and processing parts. Therefore, it can be seen that the Proposed Method utilizes the product states, worker’s two-hand processing parts, and skeleton data integratedly, and learns and evaluates taking them into account.

## V. Conclusion

We proposed a method for estimating elemental works in an assembly process consisting of multiple elemental works by integrating the product states, worker's two-hand processing parts, and worker's skeleton data using GNN. We evaluated the Proposed Method on the work movies of different 5 workers and obtained an average F-measure of 0.7833. It was found to be sufficiently accurate for estimating the elemental works.

## References

[1]Morales, J., Yoshimura, N., Xia, Q., Wada, A., Namioka, Y. and Maekawa, T. Acceleration-based Human Activity Recognition of Packaging Tasks Using Motif-guided Attention Networks, Proc. of IEEE Int'l Conf. on Pervasive Computing and Communications (PerCom 2022)

[2]Y. Sijie, X. Yuanjun, and L. Dahua, “Spatial Temporal Graph Convolutional Networks for Skeleton Based Action Recognition,” in Association for the Advancement of Artificial Intelligence, 2018.