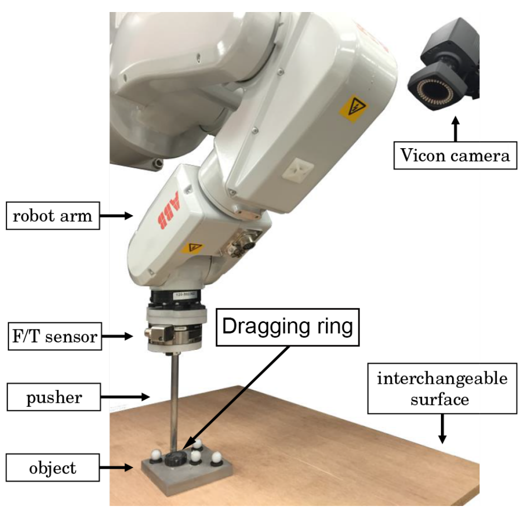
How Automated Data-Collection Dynamics Embeds Bias into Dataset

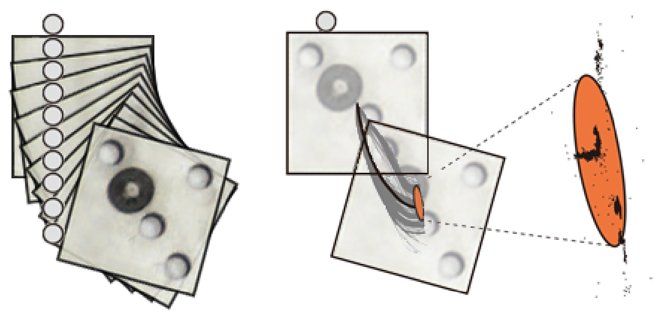
Daolin Ma[1,2] and Alberto Rodriguez[2]

[1] *Department of Civil Engineering, Southwest Jiaotong University, daolinma@mit.edu*
[2] *Department of Mechanical Engineering, Massachusetts Institute of Technology, albertor@mit.edu*

Large dataset is critical for successful implementation of machine learning. To implement data-driven approach in robotic manipulation, collecting large scale of data is no less important than designing smart algorithms. Yu [1] have collected a large dataset of repeated planar pushing to understand the frictional contacts and to prepare for data-driven manipulation. However, detailed analysis of the dataset has shown that there is severe bias embedded which resulted in structured uncertainty that was difficult to explain. The bias can be explained by analyzing the the experiment dynamics, but the frictional forces governing the objects motion should be modelled with anisotropic friction law rather than by adding random uncertainty to Coulomb friction coefficient [2]. The experiment setup is shown in Fig.1(a) and trajectories of CM of pushed object are shown in Fig.1(b). This paper will show numerical simulation of the whole data collection experiment instead of simulating one single pushing manipulation and address the importance of understanding dataset by analyzing experiment dynamics.



(a) Experiment setup of repeated pushing in [1]. The robot arm holds a cylinder pusher to push a rectangle object at a fixed position relative to the rectangle ((a,b) at object frame) in a straight line. After a pushing phase, the robot drags the object back to the central area of the surface through pulling on the black ring on object. Position and orientation were measured via Vicon system with high fidelity.

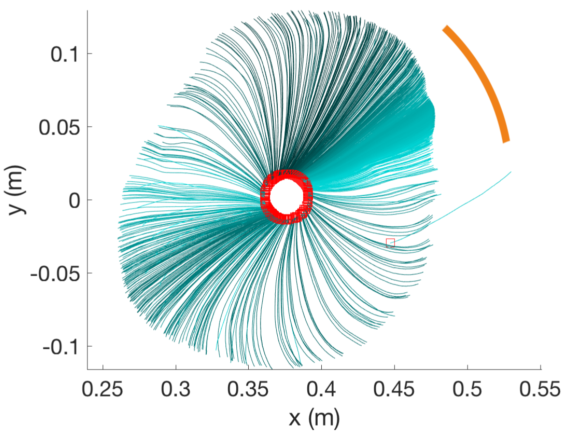


(b) Trajectories of the center of mass of an object during 2000 pushes. The figure shows an interesting distribution of end poses. This figure is adapted from [3].

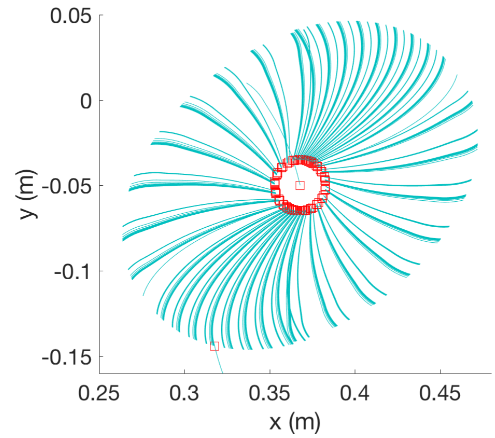
Fig. 1: Experiment setup and structured uncertainties in dataset.

Anisotropic friction is sometimes observed when studying friction between an object and another surface. This feature is usually due to micro-texture on the surface, directional orientation of molecule structure of the material or other mechanisms, which is not unnoticed apparently but has broken the isotropy in micro scale [4, 5].

We analyzed the mechanics of quasi-static pushing for a cylinder pusher and a rectangular pushed object on a rough plane with anisotropic friction. The quasi-static relationship between the instantaneous velocity of a sliding object and the support frictional force is governed by a closed, convex, origin-enclosing limit surface in (f_x, f_y, m) space representing modified from [6]. Then we investigated the whole data collection experiment dynamics by simulating sequential repeating pushing and dragging. As shown in Fig.2(a) and Fig.2(b), simulation results replicated the bias in experiment dataset. It is shown that anisotropic friction and the resetting mechanism designed to drag the object back to central area have embedded bias in collected dataset so that the initial orientation favors certain direction which might deteriorate the performance of a learned model.



(a) The trajectories of center of mass of the pushed object in experiment data.



(b) The trajectories of center of mass of the pushed object in simulation result.

Unlike accumulating data with distributed sensors where collecting is the main problem or in simulation environment where initial conditions can be precisely controlled, collecting large scale data for robotic manipulation with robots arises many challenges. For one thing, robotic manipulation involves complex frictional contacts between manipulator, targets and environment so that it's not easy to set up an experiment facility to collect data robustly. For another, the mechanism to reset initial conditions for experiment is difficult since precisely and robustly controlling the position and orientation of the target is one of research goals of robotic manipulation which is not solved yet. Let alone the high cost, wear and time-demanding of collecting data with robots. Thus it's important to analyze the experiment dynamics carefully before the data collection or before learning with the collected data.

References

[1] K.-T. Yu, M. Bauza, N. Fazeli, and A. Rodriguez, "More than a Million Ways to Be Pushed: A High-Fidelity Experimental Data Set of Planar Pushing," 2016.

[2] J. Zhou, J. A. Bagnell, and M. T. Mason, "A Fast Stochastic Contact Model for Planar Pushing and Grasping: Theory and Experimental Validation," 2017.

[3] M. Bauza and A. Rodriguez, "A Probabilistic Data-Driven Model for Planar Pushing," in *ICRA*, 2017.

[4] A. Zmitrowicz, "A theoretical model of anisotropic dry friction," *Wear*, vol. 73, no. 1, pp. 9–39, 1981.

[5] S. V. Walker and R. I. Leine, "Anisotropic dry friction with non-convex force reservoirs : modeling and experiments," vol. 1, no. c, 2017.

[6] S. Goyal, A. Ruina, and J. Papadopoulos, "Planar sliding with dry friction Part 1. Limit surface and moment function," *Wear*, vol. 143, no. 2, pp. 307–330, 1991.