

Interpretable Statistical Learning for Real-World Behavioral Data

Dissertation Defense

Patrick Emedom-Nnamdi

Harvard University

Committee: Dr. Jukka-Pekka Onnela, Dr. Junwei Lu, and Dr. Susan Murphy

May 1, 2023

Outline

1 Introduction

2 Overview of Dissertation Chapters:

- a **Chapter 1:** Nonparametric Additive Value Functions: Interpretable Reinforcement Learning with an Application to Surgical Recovery
- b **Chapter 2:** Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning
- c **Chapter 3:** Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder
- d **Chapter 4:** Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – *Piloting the Digital Assessment in Neuro-Oncology*

3 Conclusion

4 Acknowledgements

Transition from **Episodic** to Continuous Data Collection

Traditional approaches for evaluating subject behavior in healthcare rely on tools such as **surveys** or **in-clinic interviews**

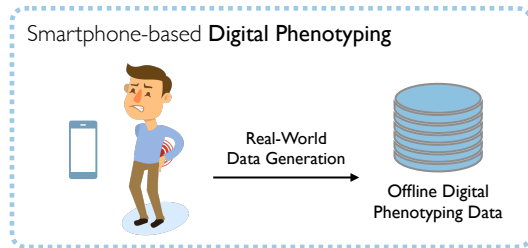
- Administered in fixed periods of time, typically after the event or episode of interest has occurred
 - ▶ Resulting in data collection that is **episodic** and **sparse** in nature
- Attempt to build a **chronological ordering of events** based on subject recall
- Build categorizations of patient experience that are heavily **contextualized** to the specific time and location of data collection



Transition from Episodic to **Continuous** Data Collection

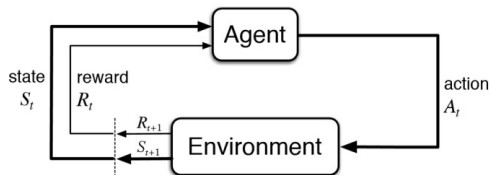
Novel data collection approaches attempt to characterize **real-world** human behavior in an **objective, consistent** manner

- Focus on identifying social, behavioral, and cognitive phenotypes that are temporally and contextually dependent
- Aim to improve **reliability** of data collected by minimizing reliance on subject recall
- Leverage devices that have (1) **high adherence** and (2) **use** among a study populations (i.e., smartphones)
- **Digital Phenotyping** is defined as “moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices, in particular **smartphones**.”

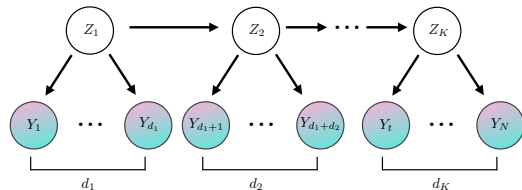


Research Overview

Goal of Dissertation: Develop interpretable statistical machine learning methodology that reveal valuable clinical insight from real-world behavioral data.



(a) (Offline) Reinforcement Learning



(b) Latent State-Space Modeling

Smartphone Data Collection and Analysis

Beiwe Data Collection: Patients installed the data collection application to their smartphones.

- Beiwe is a high-throughput raw data collection platform in development and use since 2013
- Front-end: native Android and iOS applications for collecting active and passive data
- Back-end: Amazon Web Services (AWS) cloud computing, scalable, globally deployable

Forest data analytics library provides a suite of methods for handling data generated from Beiwe

- GPS imputation and summarization, activity recognition (e.g., step counting, walking cadence), survey data aggregation, call/text summarization, etc.
- Implemented in Python and can be run locally or on the AWS backend



Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

Postoperative Spine Disease Recovery

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

Postoperative recovery is defined as the period of functional improvement between the end of surgery to the onset of normal functional activity:

- Can vary drastically among patients
- Accompanied by mild to severe complications

Current Physician-guided Recommendations:

- Early mobilization activities (e.g., as standing, walking, and getting in and out bed)
- Consistent pain management

Physicians currently lack:

- A quantification of allowable levels of mobilization (e.g., relative step count, walking cadence)
- An understanding of optimal time to initiate mobilization over the course of the recovery period



Challenges in Current Clinical Practice

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

Current measurements for postoperative assessments rely on **patient-reported outcome measures (PROMs)**

- A set of clinically-validated questionnaires that capture patient outcome
- Administered in person during follow-up visits, or, electronics via surveys
- PROMs are insufficient for most tasks
 - ▶ Underutilized by physicians due to administration burden
 - ▶ Reliant on patient recall
 - ▶ Sparse coverage over the course of patient follow-up
 - ▶ Response (shift) bias

SRS-22 Patient Questionnaire

Instructions:

- Use a #2 pencil to mark.
- Not answers must be entered with the web interface.
- All questions must be answered unless otherwise indicated.
- Complete 30 or more to receive answers.

Demographics: Last name, First name, Gender, Street, City, State, Zip, Country, Email, Telephone, Occupation, Birthdate (DD MM YYYY), Telephone.

Examination date: Day, Month, Year.

Examination interval after surgery:

Examination interval after surgery	Year
< 1 month	< 1 year
1 - 3 months	1 - 3 years
3 - 6 months	3 - 6 years
6 - 12 months	6 - 12 years
12 - 24 months	12 - 24 years
24 - 36 months	24 - 36 years
36 - 48 months	36 - 48 years
48 - 60 months	48 - 60 years
60 - 72 months	72 - 84 years
72 - 84 months	84 - 96 years
84 - 96 months	96 - 108 years
108 - 120 months	120 - 132 years
132 - 144 months	144 - 156 years
156 - 168 months	168 - 180 years
180 - 192 months	192 - 204 years
192 - 204 months	204 - 216 years
216 - 228 months	228 - 240 years
240 - 252 months	252 - 264 years
264 - 276 months	276 - 288 years
288 - 300 months	300 - 312 years
312 - 324 months	324 - 336 years
336 - 348 months	348 - 360 years
360 - 372 months	372 - 384 years
384 - 396 months	396 - 408 years
408 - 420 months	420 - 432 years
432 - 444 months	444 - 456 years
456 - 468 months	468 - 480 years
480 - 492 months	492 - 504 years
504 - 516 months	516 - 528 years
528 - 540 months	540 - 552 years
552 - 564 months	564 - 576 years
576 - 588 months	588 - 600 years
600 - 612 months	612 - 624 years
624 - 636 months	636 - 648 years
648 - 660 months	660 - 672 years
672 - 684 months	684 - 696 years
696 - 708 months	708 - 720 years
720 - 732 months	732 - 744 years
744 - 756 months	756 - 768 years
768 - 780 months	780 - 792 years
792 - 804 months	804 - 816 years
816 - 828 months	828 - 840 years
840 - 852 months	852 - 864 years
864 - 876 months	876 - 888 years
888 - 900 months	900 - 912 years
912 - 924 months	924 - 936 years
936 - 948 months	948 - 960 years
960 - 972 months	972 - 984 years
984 - 996 months	996 - 1008 years
1008 - 1020 months	1020 - 1032 years
1032 - 1044 months	1044 - 1056 years
1056 - 1068 months	1068 - 1080 years
1080 - 1092 months	1092 - 1104 years
1104 - 1116 months	1116 - 1128 years
1128 - 1140 months	1140 - 1152 years
1152 - 1164 months	1164 - 1176 years
1176 - 1188 months	1188 - 1200 years
1200 - 1212 months	1212 - 1224 years
1224 - 1236 months	1236 - 1248 years
1248 - 1260 months	1260 - 1272 years
1272 - 1284 months	1284 - 1296 years
1296 - 1308 months	1308 - 1320 years
1320 - 1332 months	1332 - 1344 years
1344 - 1356 months	1356 - 1368 years
1368 - 1380 months	1380 - 1392 years
1392 - 1404 months	1404 - 1416 years
1416 - 1428 months	1428 - 1440 years
1440 - 1452 months	1452 - 1464 years
1464 - 1476 months	1476 - 1488 years
1488 - 1500 months	1500 - 1512 years
1512 - 1524 months	1524 - 1536 years
1536 - 1548 months	1548 - 1560 years
1560 - 1572 months	1572 - 1584 years
1584 - 1596 months	1596 - 1608 years
1608 - 1620 months	1620 - 1632 years
1632 - 1644 months	1644 - 1656 years
1656 - 1668 months	1668 - 1680 years
1680 - 1692 months	1692 - 1704 years
1704 - 1716 months	1716 - 1728 years
1728 - 1740 months	1740 - 1752 years
1752 - 1764 months	1764 - 1776 years
1776 - 1788 months	1788 - 1800 years
1800 - 1812 months	1812 - 1824 years
1824 - 1836 months	1836 - 1848 years
1848 - 1860 months	1860 - 1872 years
1872 - 1884 months	1884 - 1896 years
1896 - 1908 months	1908 - 1920 years
1920 - 1932 months	1932 - 1944 years
1944 - 1956 months	1956 - 1968 years
1968 - 1980 months	1980 - 1992 years
1992 - 2004 months	2004 - 2016 years
2016 - 2028 months	2028 - 2040 years
2040 - 2052 months	2052 - 2064 years
2064 - 2076 months	2076 - 2088 years
2088 - 2100 months	2100 - 2112 years
2112 - 2124 months	2124 - 2136 years
2136 - 2148 months	2148 - 2160 years
2160 - 2172 months	2172 - 2184 years
2184 - 2196 months	2196 - 2208 years
2208 - 2220 months	2220 - 2232 years
2232 - 2244 months	2244 - 2256 years
2256 - 2268 months	2268 - 2280 years
2280 - 2292 months	2292 - 2304 years
2304 - 2316 months	2316 - 2328 years
2328 - 2340 months	2340 - 2352 years
2352 - 2364 months	2364 - 2376 years
2376 - 2388 months	2388 - 2400 years
2400 - 2412 months	2412 - 2424 years
2424 - 2436 months	2436 - 2448 years
2448 - 2460 months	2460 - 2472 years
2472 - 2484 months	2484 - 2496 years
2496 - 2508 months	2508 - 2520 years
2520 - 2532 months	2532 - 2544 years
2544 - 2556 months	2556 - 2568 years
2568 - 2580 months	2580 - 2592 years
2592 - 2604 months	2604 - 2616 years
2616 - 2628 months	2628 - 2640 years
2640 - 2652 months	2652 - 2664 years
2664 - 2676 months	2676 - 2688 years
2688 - 2700 months	2700 - 2712 years
2712 - 2724 months	2724 - 2736 years
2736 - 2748 months	2748 - 2760 years
2760 - 2772 months	2772 - 2784 years
2784 - 2796 months	2796 - 2808 years
2808 - 2820 months	2820 - 2832 years
2832 - 2844 months	2844 - 2856 years
2856 - 2868 months	2868 - 2880 years
2880 - 2892 months	2892 - 2904 years
2904 - 2916 months	2916 - 2928 years
2928 - 2940 months	2940 - 2952 years
2952 - 2964 months	2964 - 2976 years
2976 - 2988 months	2988 - 3000 years
3000 - 3012 months	3012 - 3024 years
3024 - 3036 months	3036 - 3048 years
3048 - 3060 months	3060 - 3072 years
3072 - 3084 months	3084 - 3096 years
3096 - 3108 months	3108 - 3120 years
3120 - 3132 months	3132 - 3144 years
3144 - 3156 months	3156 - 3168 years
3168 - 3180 months	3180 - 3192 years
3192 - 3204 months	3204 - 3216 years
3216 - 3228 months	3228 - 3240 years
3240 - 3252 months	3252 - 3264 years
3264 - 3276 months	3276 - 3288 years
3288 - 3300 months	3300 - 3312 years
3312 - 3324 months	3324 - 3336 years
3336 - 3348 months	3348 - 3360 years
3360 - 3372 months	3372 - 3384 years
3384 - 3396 months	3396 - 3408 years
3408 - 3420 months	3420 - 3432 years
3432 - 3444 months	3444 - 3456 years
3456 - 3468 months	3468 - 3480 years
3480 - 3492 months	3492 - 3504 years
3504 - 3516 months	3516 - 3528 years
3528 - 3540 months	3540 - 3552 years
3552 - 3564 months	3564 - 3576 years
3576 - 3588 months	3588 - 3600 years
3600 - 3612 months	3612 - 3624 years
3624 - 3636 months	3636 - 3648 years
3648 - 3660 months	3660 - 3672 years
3672 - 3684 months	3684 - 3696 years
3696 - 3708 months	3708 - 3720 years
3720 - 3732 months	3732 - 3744 years
3744 - 3756 months	3756 - 3768 years
3768 - 3780 months	3780 - 3792 years
3792 - 3804 months	3804 - 3816 years
3816 - 3828 months	3828 - 3840 years
3840 - 3852 months	3852 - 3864 years
3864 - 3876 months	3876 - 3888 years
3888 - 3900 months	3900 - 3912 years
3912 - 3924 months	3924 - 3936 years
3936 - 3948 months	3948 - 3960 years
3960 - 3972 months	3972 - 3984 years
3984 - 3996 months	3996 - 4008 years
4008 - 4020 months	4020 - 4032 years
4032 - 4044 months	4044 - 4056 years
4056 - 4068 months	4068 - 4080 years
4080 - 4092 months	4092 - 4104 years
4104 - 4116 months	4116 - 4128 years
4128 - 4140 months	4140 - 4152 years
4152 - 4164 months	4164 - 4176 years
4176 - 4188 months	4188 - 4200 years
4200 - 4212 months	4212 - 4224 years
4224 - 4236 months	4236 - 4248 years
4248 - 4260 months	4260 - 4272 years
4272 - 4284 months	4284 - 4296 years
4296 - 4308 months	4308 - 4320 years
4320 - 4332 months	4332 - 4344 years
4344 - 4356 months	4356 - 4368 years
4368 - 4380 months	4380 - 4392 years
4392 - 4404 months	4404 - 4416 years
4416 - 4428 months	4428 - 4440 years
4440 - 4452 months	4452 - 4464 years
4464 - 4476 months	4476 - 4488 years
4488 - 4500 months	4500 - 4512 years
4512 - 4524 months	4524 - 4536 years
4536 - 4548 months	4548 - 4560 years
4560 - 4572 months	4572 - 4584 years
4584 - 4596 months	4596 - 4608 years
4608 - 4620 months	4620 - 4632 years
4632 - 4644 months	4644 - 4656 years
4656 - 4668 months	4668 - 4680 years
4680 - 4692 months	4692 - 4704 years
4704 - 4716 months	4716 - 4728 years
4728 - 4740 months	4740 - 4752 years
4752 - 4764 months	4764 - 4776 years
4776 - 4788 months	4788 - 4800 years
4800 - 4812 months	4812 - 4824 years
4824 - 4836 months	4836 - 4848 years
4848 - 4860 months	4860 - 4872 years
4872 - 4884 months	4884 - 4896 years
4896 - 4908 months	4908 - 4920 years
4920 - 4932 months	4932 - 4944 years
4944 - 4956 months	4956 - 4968 years
4968 - 4980 months	4980 - 4992 years
4992 - 5004 months	5004 - 5016 years
5016 - 5028 months	5028 - 5040 years
5040 - 5052 months	5052 - 5064 years
5064 - 5076 months	5076 - 5088 years
5088 - 5100 months	5100 - 5112 years
5112 - 5124 months	5124 - 5136 years
5136 - 5148 months	5148 - 5160 years
5160 - 5172 months	5172 - 5184 years
5184 - 5196 months	5196 - 5208 years
5208 - 5220 months	5220 - 5232 years
5232 - 5244 months	5244 - 5256 years
5256 - 5268 months	5268 - 5280 years
5280 - 5292 months	5292 - 5304 years
5304 - 5316 months	5316 - 5328 years
5328 - 5340 months	5340 - 5352 years
5352 - 5364 months	5364 - 5376 years
5376 - 5388 months	5388 - 5400 years
5400 - 5412 months	5412 - 5424 years
5424 - 5436 months	5436 - 5448 years
5448 - 5460 months	5460 - 5472 years
5472 - 5484 months	5484 - 5496 years
5496 - 5508 months	5508 - 5520 years
5520 - 5532 months	5532 - 5544 years
5544 - 5556 months	5556 - 5568 years
5568 - 5580 months	5580 - 5592 years
5592 - 5604 months	5604 - 5616 years
5616 - 5628 months	5628 - 5640 years
5640 - 5652 months	5652 - 5664 years
5664 - 5676 months	5676 - 5688 years
5688 - 5700 months	5700 - 5712 years
5712 - 5724 months	5724 - 5736 years
5736 - 5748 months	5748 - 5760 years
5760 - 5772 months	5772 - 5784 years
5784 - 5796 months	5796 - 5808 years
5808 - 5820 months	5820 - 5832 years
5832 - 5844 months	5844 - 5856 years
5856 - 5868 months	5868 - 5880 years
5880 - 5892 months	5892 - 5904 years
5904 - 5916 months	5916 - 5928 years
5928 - 5940 months	5940 - 5952 years
5952 - 5964 months	5964 - 5976 years
5976 - 5988 months	5988 - 6000 years
6000 - 6012 months	6012 - 6024 years
6024 - 6036 months	6036 - 6048 years
6048 - 6060 months	6060 - 6072 years
6072 - 6084 months	6084 - 6096 years
6096 - 6108 months	6108 - 6120 years
6120 - 6132 months	6132 - 6144 years
6144 - 6156 months	6156 - 6168 years
6168 - 6180 months	6180 - 6192 years
6192 - 6204 months	6204 - 6216 years
6216 - 6228 months	6228 - 6240 years
6240 - 6252 months	6252 - 6264 years
6264 - 6276 months	6276 - 6288 years
6288 - 6300 months	6300 - 6312 years
6312 - 6324 months	6324 - 6336 years
6336 - 6348 months	6348 - 6360 years
6360 - 6372 months	6372 - 6384 years
6384 - 6396 months	6396 - 6408 years
6408 - 6420 months	6420 - 6432 years
6432 - 6444 months	6444 - 6456 years
6456 - 6468 months	6468 - 6480 years
6480 - 6492 months	6492 - 6504 years
6504 - 6516 months	6516 - 6528 years
6528 - 6540 months	6540 - 6552 years
6552 - 6564 months	6564 - 6576 years
6576 - 6588 months	6588 - 6600 years
6600 - 6612 months	6612 - 6624 years
6624 - 6636 months	6636 - 6648 years
6648 - 6660 months	6660 - 6672 years
6672 - 6684 months	6684 - 6696 years
6696 - 6708 months	6708 - 6720 years
6720 - 6732 months	6732 - 6744 years
6744 - 6756 months	6756 - 6768 years
6768 - 6780 months	6780 - 6792 years
6792 - 6804 months	6804 - 6816 years
6816 - 6828 months	6828 - 6840 years
6840 - 6852 months	6852 - 6864 years
6864 - 6876 months	6876 - 6888 years
6888 - 6900 months	6900 - 6912 years
6912 - 6924 months	6924 - 6936 years
6936 - 6948 months	6948 - 6960 years
6960 - 6972 months	6972 - 6984 years
6984 - 6996 months	6996 - 7008 years
7008 - 7020 months	7020 - 7032 years
7032 - 7044 months	7044 - 7056 years
7056 - 7068 months	7068 - 7080 years
7080 - 7092 months	7092 - 7104 years
7104 - 7116 months	7116 - 7128 years
7128 - 7140 months	7140 - 7152 years
7152 - 7164 months	7164 - 7176 years
7176 - 7188 months	7188 - 7200 years
7200 - 7212 months	7212 - 7224 years
7224 - 7236 months	7236 - 7248 years
7248 - 7260 months	7260 - 7272 years
7272 - 7284 months	7284 - 7296 years
7296 - 7308 months	7308 - 7320 years
7320 - 7332 months	7332 - 7344 years
7344 - 7356 months	7356 - 7368 years
7368 - 7380 months	7380 - 7392 years
7392 - 7404 months	7404 - 7416 years
7416 - 7428 months	7428 - 7440 years
7440 - 7452 months	7452 - 7464 years
7464 - 7476 months	7476 - 7488 years
7488 - 7500 months	7500 - 7512 years
7512 - 7524 months	7524 - 7536 years
7536 - 7548 months	7548 - 7560 years
7560 - 7572 months	7572 - 7584 years
7584 - 7596 months	7596 - 7608 years
7608 - 7620 months	7620 - 7632 years
7632 - 7644 months	7644 - 7656 years
7656 - 7668 months	7668 - 7680 years
7680 - 7692 months	7692 - 7704 years
7704 - 7716 months	7716 - 7728 years
7728 - 7740 months	7740 - 7752 years
7752 - 7764 months	7764 - 7776 years
7776 - 7788 months	7788 - 7800 years
7800 - 7812 months	7812 - 7824 years
7824 - 7836 months	7836 - 7848 years
7848 - 7860 months	7860 - 7872 years
7872 - 7884 months	7884 - 7896 years

Brigham and Women Hospital's Spine Disease Study Cohort

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

Computational Neuroscience Outcomes Center (CNOC) at Harvard and BWH constructed the first smartphone-based digital phenotyping spine disease study cohort

- **Study Aim:** Quantify real-world patient experience using *reliable* and *objective* measures of quality of life
- $n = 344$ patients with clinically diagnosed spine disease
- Median age of 57 (IQR: 49-68)
- Enrollment period between June 2016 and March 2020, with 6-months of follow-up
- 58.6% of patient received a neurosurgical intervention



Digital Phenotyping Summary Statistics

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

Mobility-based Summary Statistics:

Distance Traveled (km)	Radius of Gyration (km)	Average flight duration (km)
Time Spent at Home (hours)	Maximum Diameter (km)	Fraction of the day spent stationary
Max. Distance from Home (km)	Num. Significant Places Visited	Time Spent Walking
Average flight length (km)	Step Count	Average Cadence

Table: Subset of GPS and accelerometer-based summary statistics of digital phenotyping.

Active Data Collection (via micro-surveys):

- Active: 0-10 Self Reported Pain level
 - ▶ Prompt: “Please rate your pain over the last 24 hours on a scale from 0 to 10, where 0 is no pain at all and 10 is the worst pain imaginable”

Evaluating Postoperative Mobility and Recovery

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

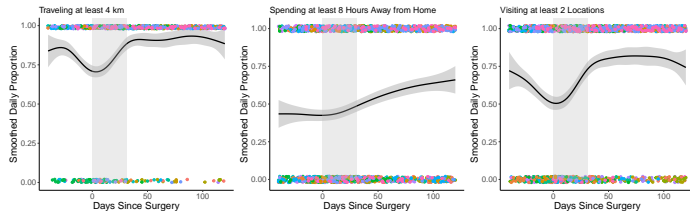


Figure: Smoothed mobility proportions (with standard errors represented in grey) for spine disease cohort centered on day of surgery.

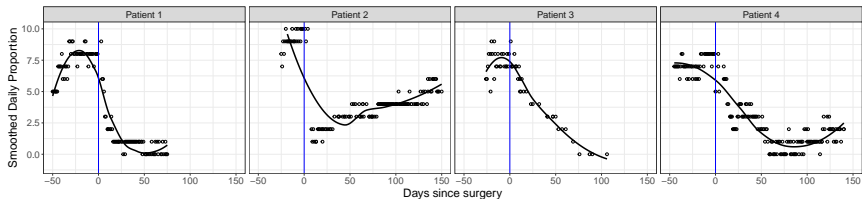


Figure: Pre- and post-operative pain responses with time centered on the day of surgery (i.e., blue line) with a fitted local regression (i.e., black line) for a random selection of patients.

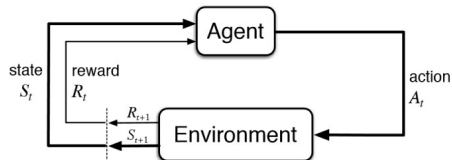
Research Problem & Objective

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

- **Problem:** How can we improve post-operative rehabilitation for surgical patients?
- **Goal:** Can we learn a decision-making policy concerning the daily post-operative steps a patient should take in order to improve their recovery process given their current condition?
- **Objective:** To estimate **interpretable recovery strategies** for post-operative surgical patients using offline digital phenotyping data

Re-framing as a Reinforcement Learning Problem

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery



Policy: is a any function $\pi : S \rightarrow A$ mapping states to actions

Q-value Function: provides a mapping from a state-action pair, (s, a) , to the *expected total discounted future rewards*

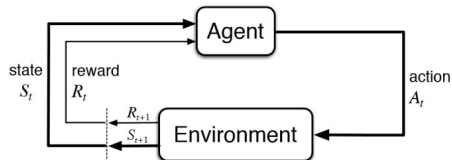
$$Q^\pi(s, a) = E \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, a_0 = a \right]$$

when following a given policy π . Note that $\gamma \in [0, 1)$

Goal is to find **optimal policy**, $\pi^*(s) = \arg \max_{a' \in \mathcal{A}} \underbrace{Q^{\pi^*}(s, a')}_{\text{learn via function approximation}}$

Need for Interpretable Function Approximators

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery



Current state-of-the-art algorithms approximate $Q^\pi(s, a)$ use **black-box methods**, e.g., neural networks

- Incorporate fitted Q-iteration with modern tools (e.g., replay buffers, target networks)
- Solve complex, high-dimensional decision-making tasks (e.g., Go, chess)

Unfortunately, the success of these modern RL algorithms comes at the **cost of model interpretability**

- Unable to examine the contribution each feature makes in producing the model's final suggested action
- Unsuitable in high-risk domains such as health care – hope to anticipate the models performance

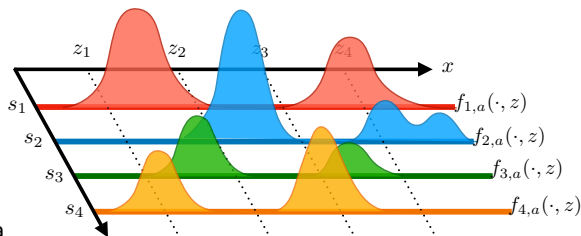
Nonparametric Additive Value Functions

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

We present a **generalized nonparametric, additive** framework for modeling Q^π :

$$Q^\pi(\mathbf{s}, a, x) = g_a(x) + \sum_{j=1}^d f_{j,a}(\mathbf{s}_j, x) + \epsilon$$

- Expand the input space of Q^π to include the variable $x \in \mathbb{R}$ (e.g., a candidate state feature, a confounder, or an action)
- $g_a(\cdot)$ represents the **additive marginal effect** of x under action a
- $f_{j,a}(\cdot, \cdot)$ represents the **additive joint effect of interactions** between x and state features \mathbf{s}_j under action a



Our model structure allows us to examine **additive nonlinear relationships** that may exist among relevant state features, actions and the variable x

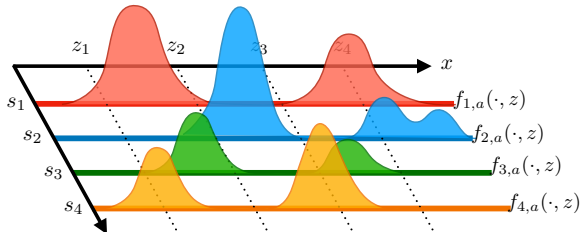
Estimation Strategy for $g_a(x)$ and $f_{j,a}(\mathbf{s}_j, x)$

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

Step 1 – Basis Expansion: For an arbitrary fixed constant z , we locally express our model using a centered B-spline basis expansion:

$$Q^\pi(\mathbf{s}, a, x = z) \approx \alpha_{a,z} + \sum_{j=1}^d \sum_{\ell=1}^m \varphi_{j\ell}(\mathbf{s}_j) \beta_{j\ell;a,z}$$

- $\alpha_{a,z}$ is a constant representing the marginal effect $g_a(\cdot)$ at the fixed value z
- Each additive component function $f_{j,a}$ is represented as a centered B-Spline basis function



$$f_{j,a}(\mathbf{s}_j, z) = \sum_{\ell=1}^m \varphi_{j\ell}(\mathbf{s}_j) \beta_{j\ell;a,z},$$

where $\varphi_{j\ell}(\mathbf{s}) = \psi_{\ell}(\mathbf{s}) - E[\psi_{\ell}(\mathbf{s}_j)]$ for the j -th component function and the ℓ -th basis component

Estimation Strategy for $g_a(x)$ and $f_{j,a}(\mathbf{s}_j, x)$

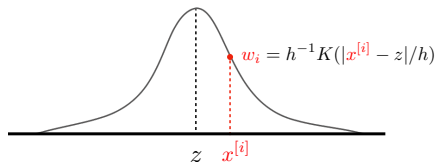
Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

Step 2 – Kernel (or Locally) Weighted Least Squares Fixed Point Approximation: Finding a good approximation of Q^π equates to forcing the approximate value function to be a fixed point under the Bellman operator \mathcal{T}_π :

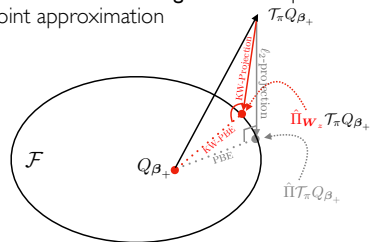
- **Classical Estimation of Q^π via LSTDQ (Lagoudakis and Parr 2004):** $Q_{\beta_+} \approx \Pi \mathcal{T}_\pi Q_{\beta_+}$
- **Kernel Weighted Approach:** $Q_{\beta_+} \approx \Pi \mathbf{W}_z \mathcal{T}_\pi Q_{\beta_+}$ (with a group Lasso penalty)

Step a. Construct kernel-weight matrix

$$\mathbf{W}_z = \text{diag}(K_h(x^{[1]} - z) \cdots K_h(x^{[N]} - z)) \in \mathbb{R}^{N \times N}$$



Step b. Perform kernel-weighted least squares fixed point approximation



Simulation Study

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

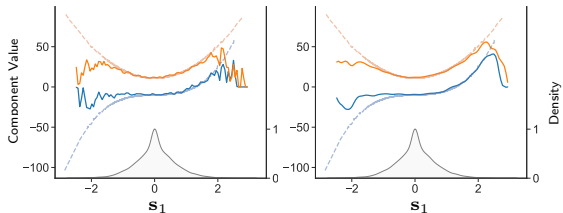
Using the simulated MDP, we examine our algorithm's performance in estimating the marginal nonlinear additive function $g_a(x)$

- We estimate the following nonparametric additive model

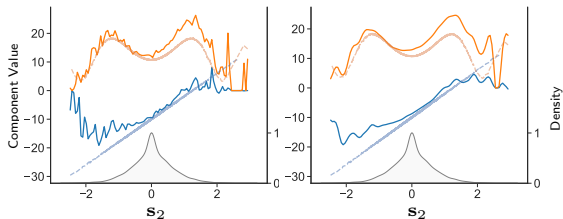
$$Q^\pi(\mathbf{s}, a, \mathbf{s}_i) = g_a(\mathbf{s}_i) + \sum_{j \in [d]/i} f_{j,a}(\mathbf{s}_j, \mathbf{s}_i) + \epsilon,$$

where $x \equiv \mathbf{s}_i$ and $i \in \{1, 2\}$

- We observe that $g_a(\mathbf{s}_i)$ can estimate the true underlying additive component $U_i(\mathbf{s}_i, a)$
 - ▶ Performs well primarily for state components that are within the observed distribution



(a) Marginal effect $\hat{g}_a(\mathbf{s}_1)$ vs. Monte-Carlo estimate of $U_1(\mathbf{s}_1, a)$.



(b) Marginal effect $\hat{g}_a(\mathbf{s}_2)$ vs. Monte-Carlo estimate of $U_2(\mathbf{s}_2, a)$.

Application to Surgical Recovery

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

Objective: Estimate a decision-making policy that suggests the **daily number of steps** necessary to reduce long-term ($\gamma = 0.5$) post-operative pain response.

- **Problem Setup:** We construct a simple MDP where each time step t corresponds to a day since surgery
 - **State vector** $s^{(t)} \in \mathbb{R}^d$ represents $d = 9$ relevant *digital phenotyping features* and patient-specific *clinical information*
 - **Actions** $a^{(t)} \in \{0, 1\}$ are binarized, where 0 corresponds to moving less than the patient-specific *pre-operative median number of steps* taken per day and 1 represents moving above this threshold
 - **Rewards** $r^{(t)} \equiv$ *negative self-reported pain score*
- We implement a set of nonparametric additive models to estimate action-value functions associated with
 - ① A **behavioral policy** π_b that aims to mimic decisions commonly taken by subjects, and
 - ② An **improved policy** π^* retrieved from performing approximate policy iteration on the estimated behavioral policy.

Sample Characteristics

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

We consider the first 60 days of recovery of $n = 67$ spine patients

- Patients follow-up period of less than 5 days were excluded
- A batch dataset $\mathcal{D} = \{(\mathbf{s}^{(i)}, a^{(i)}, r^{(i)}, \mathbf{s}'^{(i)})\}_{i=1}^N$ with $N = 1,409$ daily transitions was constructed
- A nonparametric additive model was estimated for each candidate variable of interest:
 - ▶ $x \equiv$ age, number of days since surgery

Variable	n (%) or Median (25 th –75 th)
Demographic Data	
Age	57.0 (48.0–65.5)
Female gender	34 (50.7)
Site of surgery	
Cervical	19 (28.4)
Lumbar	27 (40.3)
Thoracic	2 (3.0)
Multiple	18 (26.9)
Data Collection	
GPS days of follow-up	61 (49–61)
Accelerometer days of follow-up	61 (50.5–61)
Daily pain survey response rate	59.4 (42.4–76.9)
Digital Phenotypes	
Number of places visited	3 (2–5)
Time spent at home (hours)	18.3 (12.9–21.9)
Distance traveled (km)	32.3 (10.8–62.3)
Maximum distance from home (km)	10.6 (4.5–25.5)
Radius of gyration (km)	1.50 (0.18–5.01)
Time spent not moving	21.2 (20.2–22.2)
Average cadence	1.64 (1.55–1.74)
Number of steps	948.6 (356.9–2,005)

Marginal Effects $\hat{g}_a(x)$

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

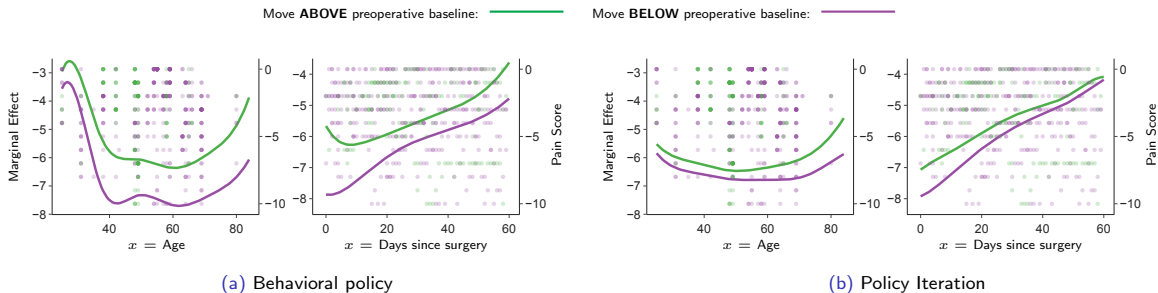
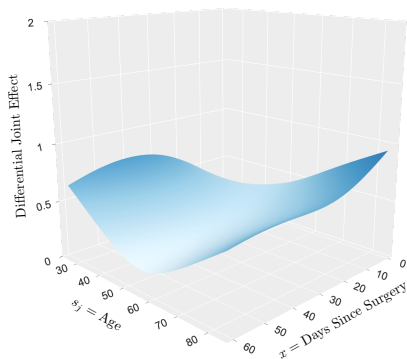


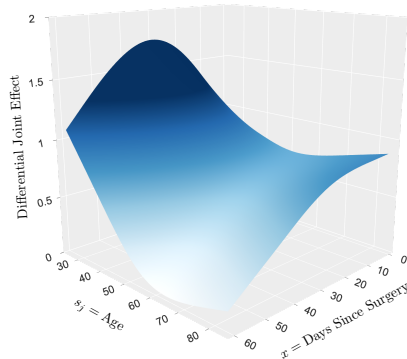
Figure: A comparison of the marginal component function $\hat{g}_a(x)$ of $Q^\pi(s, a, x)$ estimated under the behavioral policy $\pi = \pi_b$ vs. the improved policy $\pi = \pi^*$.

Differential Joint Effects $\hat{f}_{j,1}(\mathbf{s}_j, x) - \hat{f}_{j,0}(\mathbf{s}_j, x)$

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery



(a) Behavioral policy



(b) Policy Iteration

Figure: The differential benefit of selecting action $a = 1$ over $a = 0$ with respect to joint effects $\hat{f}_{j,a}(\mathbf{s}_j, x)$ under $Q^\pi(\mathbf{s}, a, x)$ estimated for the behavioral policy $\pi = \pi_b$ vs. the improved policy $\pi = \pi^*$.

Conclusion

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

- We introduce a **flexible** and **interpretable representation** for modeling action-value functions
- Our representation allows for the estimation of **non-linear marginal effect** of select variables and **joint effects** between state features
- Our modeling approach can also accommodate **continuous actions** (i.e., $x \equiv a$)
- In our application to surgical recovery, we reveal **recovery strategies** that are in-line with current clinical practice

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

Motivation

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

In most real-world applications of RL,

- Continual improvement of decision-making policies during environmental interaction is **infeasible** due to **resource/safety constraints**
- Policies are learned in an offline or a **growing-batch** manner
- Learning optimal policies is difficult due to limited opportunities for **online self-corrections**

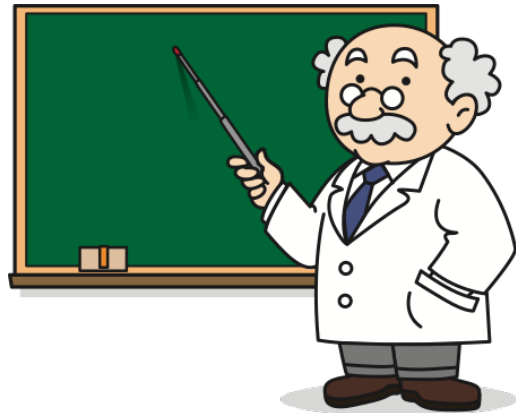


Motivation

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

In many domains, **well-informed/task-specific knowledge** exists and can be queried

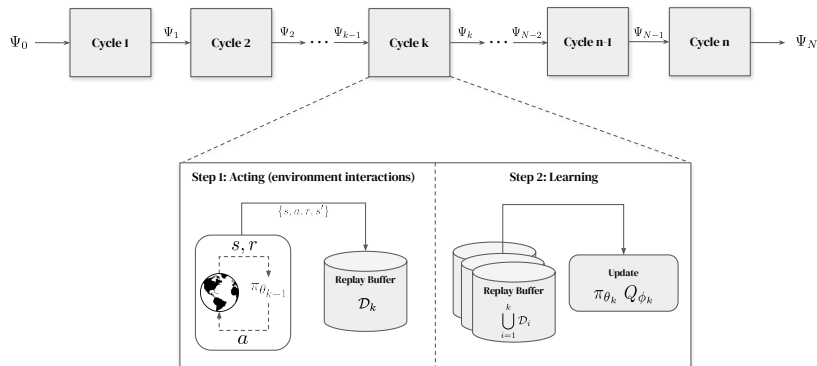
- An external teacher can help alleviate some of these issues, providing **external corrective feedback**
- Teachers can be represented as **human, RL agent, or program**
- Expert knowledge by the teacher can be provided
 - ① Up-front via **demonstrations**, or
 - ② Throughout the life-cycle of the student agent via **annotations**



How can we leverage teachers to improve the sample efficiency and performance of value-based RL agents learned in the growing-batch setting?

Growing-batch Reinforcement Learning

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning



Policy improvements via offline parameter updates are made only after new batches of experiential data are gathered from the environment.

- In practice, the number of cycles tend to be small, while the size of each newly gathered batch is large

Experimental Setup - Actor Critic Agents

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

For our experiments, we primarily utilize **D4PG agents** to accommodate continuous actions, where we

- 1 Learn a critic $Q_\phi(s, a)$ via **distributional TD-Learning**, where $Q_\phi(s, a) = \mathbb{E}Z_\phi(s, a)$ and $Z_\phi(s, a)$ is estimated by minimizing the distributional TD error

$$\mathcal{L}(\phi) = \mathbb{E}_{s \sim \rho^\pi} \left[d \left(\mathcal{T}_{\pi_\theta}, Z_{\phi'}(s, a), Z_\phi(s, a) \right) \right]$$

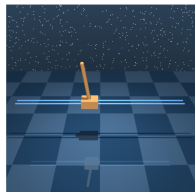
- 2 Learn a parametric policy π_θ using the **deterministic policy gradient (DPG)**

$$\nabla_\theta \mathcal{J}(\theta) = \mathbb{E}_{s \sim \rho^\pi} \left[\nabla_\theta \pi_\theta(s) \nabla_a Q_\phi(s, a) \Big|_{a=\pi_\theta(s)} \right]$$

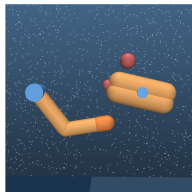
where $\mathcal{J}(\theta) = \mathbb{E}_{(s,a) \sim \mathcal{D}} [Q^{\pi_\theta}(s, a)]$ is the expected return.

Experimental Setup - DeepMind Control Suite

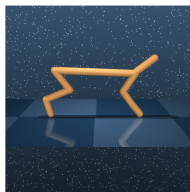
Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning



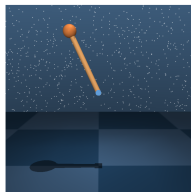
(a) Cart-pole: Balance & Swing-up



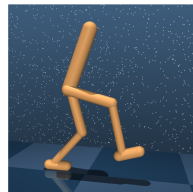
(b) Finger-Spin



(c) Cheetah-Run



(d) Pendulum Swing-up



(e) Walker-Run

Figure: The DeepMind Control Suite environments used in our experiments.

Growing-Batch and Environment Settings:

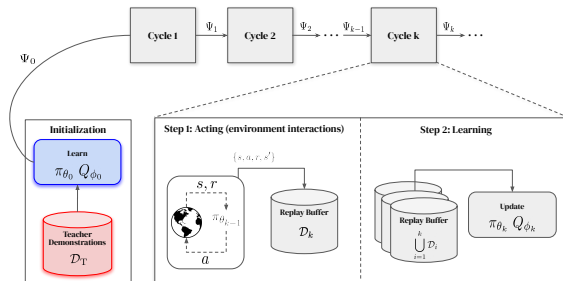
- Evaluated experiments on **six control suite environments**
- Fixed the **total number of actor steps** to 2M, which is the value needed to solve the task by an online agent

Pre-training Agents using Demonstrations

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

In our first set of experiments, we examine the performance of **naively pre-training a policy** using demonstrations from a teacher performing the task

- Pre-loaded replay buffer with **1M transitions** from 1K episodes
- Pretrained the student's policy π_{θ_0} using **behavioral cloning**
- Using the BC policy, the **critic** is pre-trained using D4PG **policy evaluation**

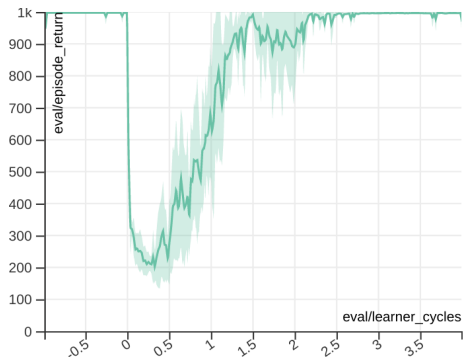


Pitfalls

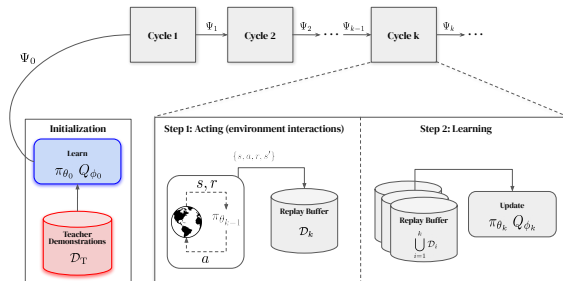
Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

Naively pretraining π_0 using **behavioral cloning performs poorly**

- **Dramatic drop in performance** when training with newly generated data **within the first cycle**



(a) Cartpole Balance, Number of Cycles = 4



BC-Policy Regularization

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

To prevent “forgetting” the initialized policy, we explored incorporating a **regularization term** on the DPG objective function:

$$\mathcal{J}(\theta_k) = \mathcal{J}_{\text{DPG}}(\theta_k) + \underbrace{\mathbb{E}_{s \sim \rho^\pi} \left[\|\pi_{\text{BC}}(s) - \pi_{\theta_k}(s)\|_2^2 \right]}_{\text{BC regularizer}}$$

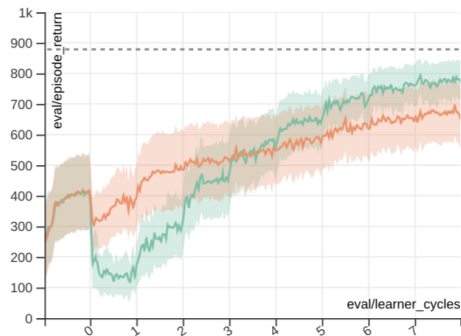
where we explicitly keep our learned policy **close to the BC-initialized policy**.

Results - BC-Policy Regularization

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning



(a) All Environments, Number of Cycles = 4



(b) All Environments, Number of cycles = 8

Figure: Baseline with BC initialization only vs. BC-policy regularizer

BC-Policy Regularization (Blissful Ignorance)

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

Assuming that the BC-initialized policy is **sub-optimal**, we would like the agent to **surpass it's performance**:

- Incorporated an **exponential decay weight** $\alpha \in (0, 1)$ as a hyper-parameter

$$\mathcal{J}(\theta_k) = (1 - \alpha) \mathcal{J}_{\text{D4PG}}(\theta_k) + \underbrace{\alpha \mathbb{E}_{s \sim \rho^\pi} \left[\left\| \pi_{\text{BC}}(s) - \pi_{\theta_k}(s) \right\|_2^2 \right]}_{\text{BC regularizer}}$$

- By treating α as a **function of total learner steps**,
 - ① We first **stay close** to the initialized policy
 - ② Then, gradually **prioritize solely learning from the DPG** loss component

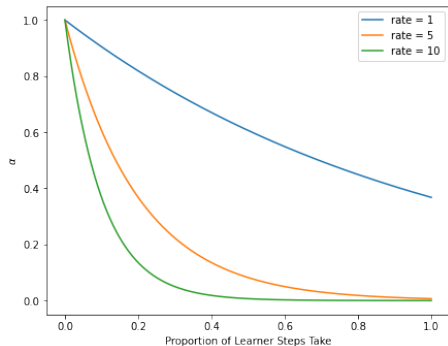


Figure: Exponential decay weight α for various rates.

BC-Policy Regularization (Blissful Ignorance)

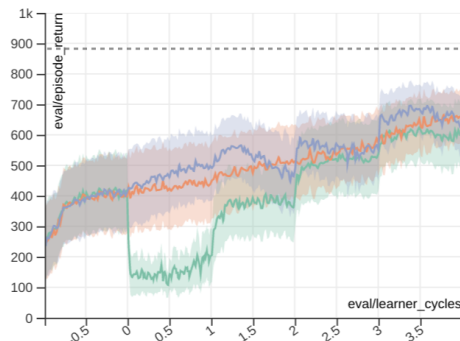
Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

Assuming that the BC-initialized policy is **sub-optimal**, we would like the agent to **surpass it's performance**:

- Incorporated an **exponential decay weight** $\alpha \in (0, 1)$ as a hyper-parameter

$$\mathcal{J}(\theta_k) = (1 - \alpha) \mathcal{J}_{\text{D4PG}}(\theta_k) + \underbrace{\alpha \mathbb{E}_{s \sim \rho^\pi} \left[\left\| \pi_{\text{BC}}(s) - \pi_{\theta_k}(s) \right\|_2^2 \right]}_{\text{BC regularizer}}$$

- By treating α as a **function of total learner steps**,
 - ① We first **stay close** to the initialized policy
 - ② Then, gradually **prioritize solely learning from the DPG** loss component



(a) All Environments, Number of cycles = 4

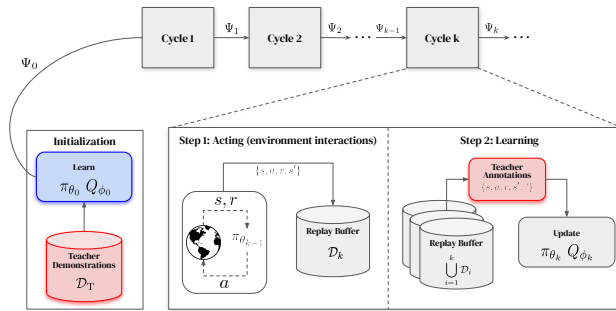
Figure: Baseline with **BC initialization only** vs. BC-policy regularizer with **decay rate = 1** and **decay rate = 5**.

Teacher Guided Annotations

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

While BC provides an initialization to a good policy, our agent still risks learning a sub-optimal policy due to the following issues related to (1) **insufficient state-action coverage** and (2) **overestimation bias**:

- We considered various forms of **annotations** provided by a teacher at training time
- We explore using **teacher-suggested actions** as a mechanism for mimicking the teacher



Teacher-Suggested Actions - Varying Decay Rates

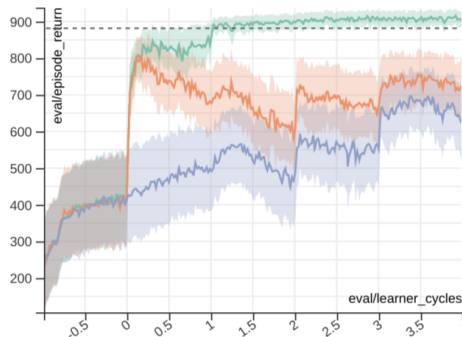
Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

We query the teacher's action $a^* \sim \pi^*(s)$ during training time:

- We incorporate an ℓ_2 loss component that evaluates the difference between **the student's policy** and **the teacher-suggested action**
- Additionally, we incorporate a **between-cycle regularizer** to promote monotonic improvement between cycles

$$\begin{aligned} \mathcal{J}(\theta_k) = & (1 - \alpha) \mathcal{J}_{\text{D4PG}}(\theta_k) \\ & + \underbrace{\beta_k \mathbb{E}_{s \sim \rho^\pi} \left[\|\pi_{\theta_{k-1}}(s) - \pi_{\theta_k}(s)\|_2^2 \right]}_{\text{Between-cycle policy regularizer}} \\ & + \underbrace{\alpha \mathbb{E}_{s \sim \rho^\pi} \left[\|a^* - \pi_{\theta_k}(s)\|_2^2 \right]}_{\text{Teacher-action Component}} \end{aligned}$$

- The weight $\beta_k \in (0, 1)$ decays with respect to cycle k



(a) All Environments, Number of Cycles = 4

Figure: Baseline using BC-policy regularizer vs. Teacher-action auxiliary loss with decay rate = 1, and decay rate = 5.

Filtered Teacher-Suggested Actions

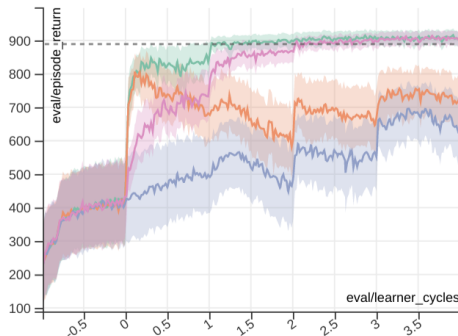
Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

We explore using a Q-filter to **determine which loss component (Teacher-action vs. D4PG) to learn from** for a given transition:

$$\nabla \mathcal{J}(\theta_k) = \underbrace{\mathbb{E}_{s \sim \rho^\pi} \left[[1 - \delta(s)] \nabla_{\theta_k} \pi_{\theta_k} \nabla_a Q_{\phi_k}(s, a) \right]_{a=\pi_{\theta_k}(s)}}_{\text{D4PG component}} + \underbrace{\delta(s) \nabla_{\theta_k} \|a^* - \pi_{\theta_k}(s)\|_2^2}_{\text{Teacher-action component}},$$

where $\delta(s) = 1 \left[Q_{\phi_k}(s, a^*) \geq Q_{\phi_k}(s, \pi_{\theta_k}(s)) \right]$ is a **Q-filter**.

- Ignore (or de-prioritize) learning from actions that produce lower values than using $\pi_{\theta_k}(s)$



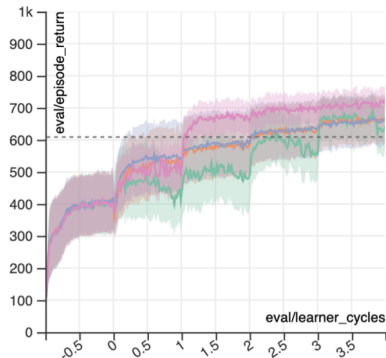
(a) All Environments, Number of Cycles = 4

Figure: Baseline using BC-policy regularizer vs. **Filtered Teacher-action** vs. Teacher-action auxiliary loss with **decay rate = 1**, and **decay rate = 5**.

Leveraging Sub-optimal Teachers

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

Information from sub-optimal teachers can be used to “jump-start” new policies:



(a) All Environments, Number of Cycles = 4

Figure: Baseline using BC-policy regularizer vs. Filtered Teacher-Actions vs. Teacher-Gradient auxiliary loss with decay rate = 1, and decay rate = 5.

Alternative Annotations - Teacher Gradients

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning

At a first glance of the **deterministic policy gradient**, one approach is to replace $\nabla_a Q_\phi(s, a)|_{a=\pi_\theta(s)}$, with the **gradient of the teacher's Q-function** w.r.t to the students policy:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_{s \sim \rho^\pi} \left[\nabla_\theta \pi_\theta(s) \nabla_a Q^{\pi^{\text{teacher}}}(s, a) \Big|_{a=\pi_\theta(s)} \right]$$

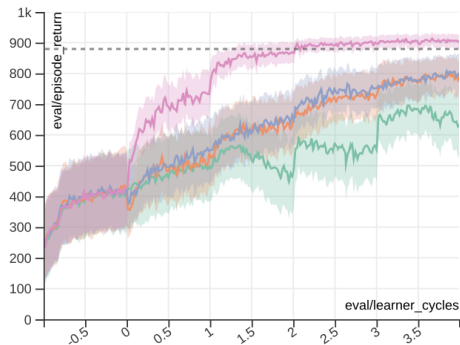
In practice, this can be achieved by modifying the DDPG policy loss as follows:

$$J(\theta_k) = (1 - \alpha) J_{\text{DDPG}}(\theta_k) + \underbrace{\alpha \|\nabla_a Q^{\pi^{\text{teacher}}}(s, a) + a - \pi_{\theta_k}(s)\|_2}_{\text{Expert Gradient DPG Component}} + \underbrace{\mathbb{E}_{s \sim \rho^\pi} [\|\pi_{\text{BC}}(s) - \pi_{\theta_k}(s)\|_2^2]}_{\text{BC regularizer}}$$

- **Intuition:** This loss encourages the agent to learn a policy that selects actions in the direction that maximizes the teacher's critic.

Results - Teacher Gradients

Chapter 2 - Knowledge Transfer from Teachers to Learners in Growing-Batch Reinforcement Learning



(a) All Environments, Number of Cycles = 4

Figure: Baseline using BC-policy regularizer vs. Filtered DAgger vs. Expert Gradient auxiliary loss with decay rate = 1, and decay rate = 5.

Conclusion

- ① Initially staying close to the BC policy completely avoids the previous drop in performance, but can dampen overall performance if “forgotten” slowly.
- ② Learning directly from the teacher's action provides a substantial performance boost, but is **sensitive to the choice rate parameter** for α .
- ③ While the Q-filter initially under-performs, it **adaptively** reaches superior performance **without** use of hyper-parameter tuning.

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

Motivation

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

Bipolar disorder is associated with significant morbidity and mortality

- 2-3 fold **increased risk** for pre-mature mortality (e.g., cardiovascular disease, diabetes, COPD, unintentional injuries, suicide)
- NIMH estimates BP is prevalent among 2.6% of US adults and 2.9% of adolescents (Ages 13–18)
- Recurrent mood episodes are common (i.e., emotional highs or lows)
 - ▶ Risk of **future episodes** increases with each occurrence
 - ▶ The number of episodes are **associated with poor outcomes**
 - ▶ Symptoms typically emerge prior to the **onset of an episode**
- Meta-analysis (2016) revealed that energy changes, increased activity, sleep problems, and physical agitation are common prior to BP episodes

Motivation

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

Study Goal: Use digital phenotyping to develop features that can help identify the onset of notable mood episodes

Current Objective: Decode *resting* and *active* patterns of smartphone use that correlate with sleep



Study Cohort

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

We study a subset of participants enrolled at NYU Langone Health & Northwell Health

- $n = 43$ adolescents (median age of 16):
 - ▶ 20 participants had a **diagnosis of bipolar disorder** and are risk of developing a **new mood episode**
 - ▶ 23 participants were selected as **healthy controls**
- Each participant installed the Beiwe research platform on their smartphones
- Examined up to **1-year of data collected** starting from January 1, 2021

Variable	n (%) or Median (25 th –75 th)
Demographic Data	
Age	16 (15–17)
Sex	
Female	23 (53.5%)
Male	18 (41.9%)
Other	2 (4.6%)
Diagnosis	
Bipolar	22 (51.2%)
Typically-developing	21 (48.8%)
Race	
Asian	2 (5)
Asian, Black or African American	2 (5%)
Asian, Black or African American, White	1 (2%)
American Indian or Alaska Native, Asian	1 (2%)
Black or African American	3 (7%)
Black or African American, White	2 (5%)
South Asian	3 (7%)
White	29 (67%)
Data Collection	
Phone usage days of follow-up	223 (116–322)
Number of surveys recorded	98 (53–148)

Smartphone Data Collection

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

Data analyzed were primarily collected from smartphone usage logs

- We focused on events that indicate moments of smartphone interactivity:
 - ▶ **iOS devices:** the smartphone being “Unlocked”
 - ▶ **Android devices:** the smartphone’s screen being “turned on”
- Data were recorded as the event occurred in real-time
- Events were binned into **five-minute intervals** starting from the onset of the day of first data collection for patient i (i.e., 00:00:00 - 00:05:00)
- Periods within a day where no usage events were recorded received an entry of 0
- Entire days where no usage events were recorded
- We denote the number of events that occurred with the interval t as $Y_t \in [0, \infty)$
- Across all patients, we collected 2.63M observations

Hidden Semi-Markov Models

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

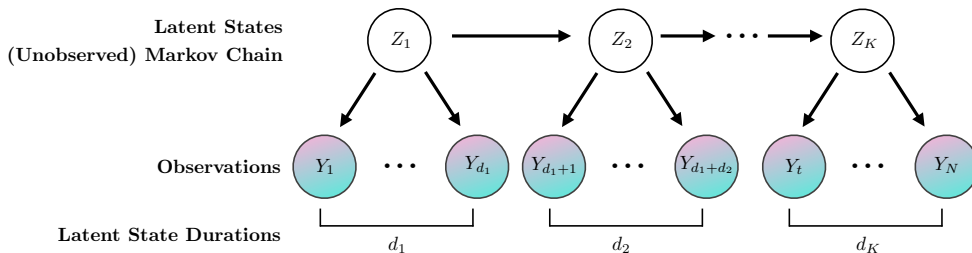
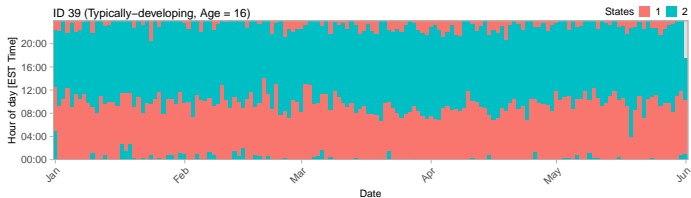


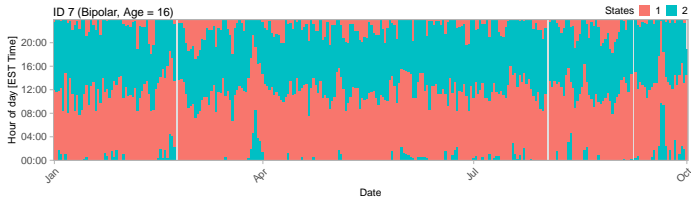
Figure: Representation of a Hidden-Semi Markov Model with state durations d_j

Decoding Viterbi Paths

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder



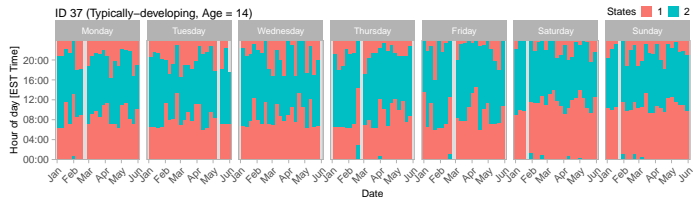
(a) Viterbi path of Participant ID 39 (*typically-developing*, age 16).



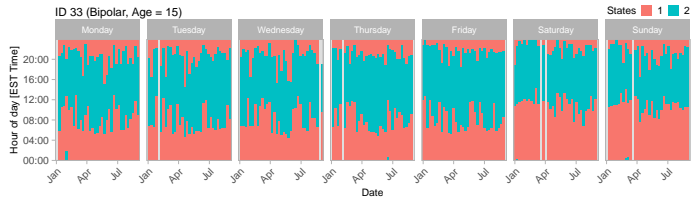
(b) Viterbi path of Participant ID 7 (*bipolar*, age 16).

Decoding Viterbi Paths – Day of the Week Effects

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder



(a) Viterbi path of Participant ID 37 (*typically-developing*, age 14) stratified by day of the week.



(b) Viterbi path of Participant ID 33 (*bipolar*, age 15) stratified by day of the week.

Latent Regression Analysis

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

Using the decoded states, we performed a regression analysis using a linear mixed effects models:

- a Model A:** To identify temporally-dependent trends between time of day and phone use under the resting and active states
- b Model B:** To identify a differential state duration between bipolar participants and healthy controls

In both model, we adjust for the same set of potential confounders: age, sex, race, and day of the week.

Model A – Temporally-dependent Phone Use

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

Model A: To identify temporally-dependent trends between time of day and phone use under the resting and active states

- We fit the following linear mixed effects model:

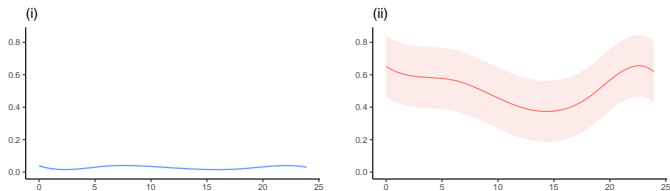
$$Y_{ti}|Z_{ti} = j \sim \mu + f(X_{ti}) + \mathbf{Z}_{ti}^{\top}\alpha + \gamma_i + \epsilon_{ti}, \quad j \in \{\text{resting, active}\}$$

where μ is the intercept, X_{ti} is our variable of interest, \mathbf{Z}_{ti} is a set of potential confounders, and γ_i is a subject-specific random intercept

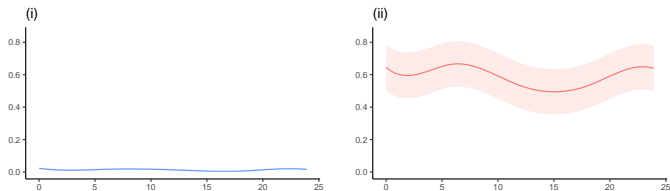
- We condition the response variable Y_{ti} on j , the value of the decoded latent state
- $X_{ti} \in [0, 24)$ represents time in hours
- f is a B-spline function with knots at 6, 12, and 18

Model A – Temporally-dependent Phone Use, $\hat{f}(X_{ti})$

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder



(a) Temporally-dependent smartphone use for *typically-developing* adolescents



(b) Temporally-dependent smartphone use for *bipolar* adolescents

Latent Regression Analysis

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

Model B: To identify a differential state duration between bipolar participants and the health controls

- We fit the following linear mixed effects model:

$$\hat{d}_{ti}|Z_{ti} = j \sim \mu + f(X_{ti}) + \mathbf{Z}_{ti}^{\top}\alpha + \gamma_i + \epsilon_{ti}, \quad j \in \{\text{resting, active}\}$$

where μ is the intercept, X_{ti} is our variable of interest, \mathbf{Z}_{ti} is a set of potential confounders, and γ_i is a subject-specific random intercept

- We condition the response variable \hat{d}_{ti} on j , the value of the decoded latent state
- \hat{d}_{ti} is the estimated state duration on day t under participant i and state j
- $f(X_{ti}) = \beta \cdot X_{ti}$, where X_{ti} is an indicator function for Bipolar status

Latent Regression Analysis

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

Model B: To identify a differential state duration between bipolar participants and the health controls

- We fit the following linear mixed effects model:

$$\hat{d}_{ti}|Z_{ti} = j \sim \mu + f(X_{ti}) + \mathbf{Z}_{ti}^{\top}\alpha + \gamma_i + \epsilon_{ti}, \quad j \in \{\text{resting, active}\}$$

where μ is the intercept, X_{ti} is our variable of interest, \mathbf{Z}_{ti} is a set of potential confounders, and γ_i is a subject-specific random intercept

- We condition the response variable \hat{d}_{ti} on j , the value of the decoded latent state
- \hat{d}_{ti} is the estimated state duration on day t under participant i and state j
- $f(X_{ti}) = \beta \cdot X_{ti}$, where X_{ti} is an indicator function for Bipolar status

Results:

Latent State	β	p-value	95%-CI
1 (Resting)	71.03	0.002	[27.64, 114.43]
2 (Active)	-69.83	0.005	[-116.34, -23.34]

Table: β estimated from LMM controlling for age, sex, race, and day of the week.

Conclusion

Chapter 3 - Decoding Resting and Active Patterns of Smartphone Use in Adolescents with Bipolar Disorder

- Decoded states provide a “first-order” approximation of sleep
- Several natural extensions:
 - ▶ Incorporate additional modalities (e.g., accelerometer/distance-metrics, screen-time) to model our observations
 - ▶ Consider heterogeneous dwell-times using covariates and/or random effects
 - ▶ Consider hierarchical hidden semi-Markov models
- **Future:** Build interventions using reinforcement learning
 - ▶ Latent states provide us with a direct and interpretable mechanism for representing observations

Chapter 4 - Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – Piloting the Digital Assessment in Neuro-Oncology

Background

Chapter 4 - Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – Piloting the Digital Assessment in Neuro-Oncology

Glioblastoma (GBM) is a rare, fast-growing, and aggressive brain tumor

- Survival is poor — 5-year survival rate $< 5\%$
- Clinical trials evaluating new treatments for GBM rely on metrics that ignore the debilitating **functional and symptomatic impact** of GBM and its **associated treatment** on patients
- Data on patient functioning and quality of life in the glioblastoma (GBM) population in free living settings are limited
 - ▶ Infrequent and sparse collection of patient-reported outcomes
- Digital phenotyping provides a means to monitor patient treatment response and recovery in an objective, consistent manner, with **minimal patient burden**

Objective: Conduct a Digital Assessment in Neuro-Oncology (DANO) pilot by identifying key trends in post-operative recovery among GBM patients

Study Cohort

Chapter 4 - Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – Piloting the Digital Assessment in Neuro-Oncology

Our study cohort consisted of 15 GBM patients (mean age 56.5; 46% female)

- 12 were **primary GBM** (no history of previous low-grade tumor)
- 3 were **recurrent/transformed** (i.e., low grade glioma transformed in higher grade)
- Each received **surgical resection** of the tumor
- A **control group** of 30 non-operative spine patients with no history of brain cancer
 - ▶ (Mean age 54.2; 33.3% female)

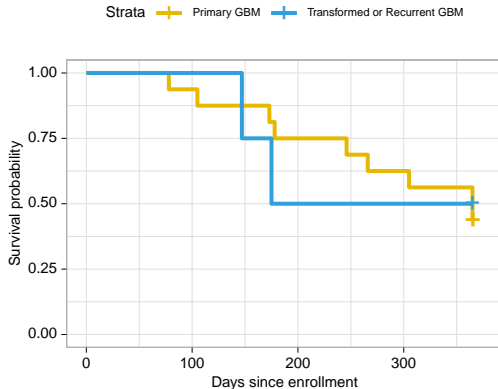


Figure: Kaplan-Meier survival curve from the start of patient enrollment to end of follow-up.

Smartphone Data Collection

Chapter 4 - Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – Piloting the Digital Assessment in Neuro-Oncology

Mobility-based Summary Statistics:

Distance Traveled (km)	Radius of Gyration (km)	Num. Significant Places Visited
Time Spent at Home (hours)	Max. Distance from Home (km)	—

Table: Subset of GPS-based summary statistics of digital phenotyping.

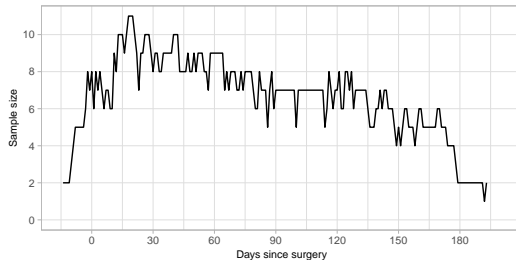
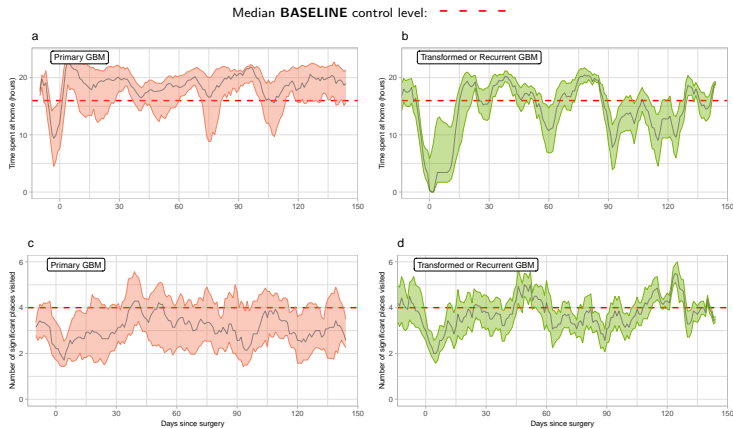


Figure: Sample size as a function of days since surgery. Number of patient-level GPS data available as a function of day since surgery. Average 108.9 days of GPS data per patient (range of 20 to 181).

Evaluating Changes in Mobility – Location-based Summary Statistics

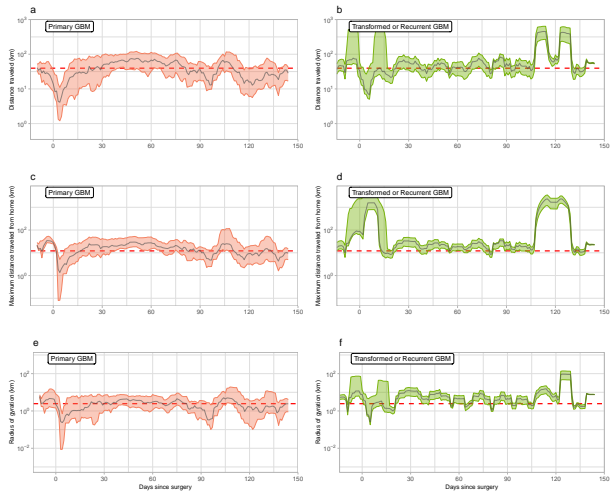
Chapter 4 - Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – Piloting the Digital Assessment in Neuro-Oncology



Evaluating Changes in Mobility – Distance-based Summary Statistics

Chapter 4 - Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – Piloting the Digital Assessment in Neuro-Oncology

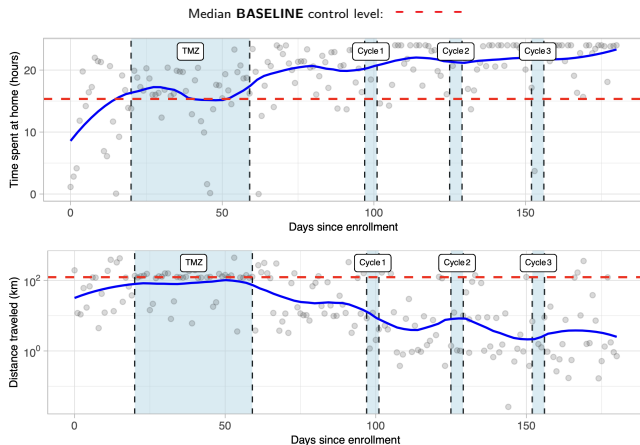
Median **BASELINE** control level: - - - -



Individual-level Changes in Mobility During Active Treatment

Chapter 4 - Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – Piloting the Digital Assessment in Neuro-Oncology

TMZ \equiv **Temozolomide**



Conclusion

Chapter 4 - Assessing Mobility in Glioblastoma Patients using Digital Phenotyping – Piloting the Digital Assessment in Neuro-Oncology

- Digital phenotyping has the potential to allow for **quantification of patient behavior and recovery** without the need for active **patient involvement**
- In general, GBM patients appear to be **less mobile** than the control group
 - ▶ With largest dip in mobility occurring immediately after surgery
- Individual-level analysis revealed noticeable dips in mobility immediately following treatment
 - ▶ Potentially attributable to **accumulating symptom burden**
- **Future** – Determine whether inferences made using mobility-based assessments coincide with **patient-reported outcomes**

Acknowledgments

Acknowledgements

Dissertation Committee:

- Dr. Jukka-Pekka Onnela
- Dr. Junwei Lu
- Dr. Susan Murphy

Collaborators:

- Dr. Timothy R. Smith
- Dr. Anna Van Meter
- Dr. Matt W. Hoffman
- Dr. Nando de Freitas
- Dr. Abram L. Friesen
- Dr. Bobak Shahriari
- Dr. Francesca Siddi
- Dr. Michael Catalino
- John L. Kilgallon
- Noah L.A. Nawabi

- Dr. Jakob V.E. Gerstl
- Jacob G. Ellen
- Krish M. Maniar

Lab Members:

- Dr. Marcin Straczekiewicz
- Dr. Debbie Huang
- Dr. Marta Karas
- Dr. Anna Beukenhorst
- Hassan Y. Dawood
- Eli Jones
- Kenzie W. Carlson
- Greyson Liu
- Zach Clement
- Octavious Smiley
- Georgios Efstathiadis

- Dr. Hali Hambridge
- Dr. Thien Le
- Dr. Till Hoffman

Biostatistics Dept.:

- Dr. Paige L. Williams
- Jelena Follweiler
- Dr. Marcello Pagano
- Dr. Bethany Hedt-Gauthier
- Dr. John Quackenbush
- Dr. Sebastien Haneuse



COMPUTATIONAL
NEUROSCIENCE
OUTCOMES CENTER
AT HARVARD



Acknowledgements



Acknowledgements



Next Steps

2023-2024 Harvard Business School Blavatnik Fellow



**Harvard
Business
School**

Blavatnik Fellowship
in Life Science
Entrepreneurship



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

raia

AI Supportive Care

Thank you!

Appendix: Simulation Study

Chapter 1: Nonparametric Additive Value Functions – Interpretable Reinforcement Learning with an Application to Surgical Recovery

We examine our algorithm's performance in estimating the marginal nonlinear additive function $g_a(x)$

- We construct a **simulated MDP**:

- **Initial state vector** $\mathbf{s}^{(0)} \in \mathbb{R}^d$ with each element sampled as $\mathbf{s}_i^{(0)} \in \text{Unif}(\frac{1}{2}, \frac{1}{2})$
- **Next state transition** occurs as $\mathbf{s}^{(t)} \sim \mathcal{N}(\mathbf{s}^{(t-1)} + \delta_a, 0.1)$, where $\delta_a = 0.1 \times \mathbb{1}(a^{(t)} = 0) - 0.1 \times \mathbb{1}(a^{(t)} = 1)$
- **Additive reward function** defined as $r(\mathbf{s}, a) = u_1(\mathbf{s}_1, a) + u_2(\mathbf{s}_2, a)$
- Perform Monte-Carlo simulation to estimate the action-value function associated with the random policy $\pi_e(s^{(t)}) \in \{0, 1\}$ with $p = 0.5$ and a discount factor of $\gamma = 0.5$

$$\begin{aligned} Q^{\pi_e}(\mathbf{s}, a) &= \mathbb{E}_{\pi_e} \left[\sum_{i=0}^{\infty} \gamma^i r(\mathbf{s}^{(i)}, a^{(i)}) \mid \mathbf{s}^{(0)} = \mathbf{s}, a^{(0)} = a \right] \\ &= \sum_{j=1}^2 \mathbb{E}_{\pi_e} \left[\sum_{i=0}^{\infty} \gamma^i u_j(\mathbf{s}_j^{(i)}, a^{(i)}) \mid \mathbf{s}^{(0)} = \mathbf{s}, a^{(0)} = a \right] \\ &= U_1(\mathbf{s}_1, a) + U_2(\mathbf{s}_2, a). \end{aligned}$$