Data Generation Tools

Team HELIOS2024

Hidehisa AKIYAMA, Tomoharu NAKASHIMA, Akei HISHIKI, Sota OKAYAMA, Shota YANO

Introduction and Background

RoboCup Soccer

- Primary objective
  - Development of soccer robots that win against human champions
- Secondary objectives
  - Development of physical robots
  - Development of decision making

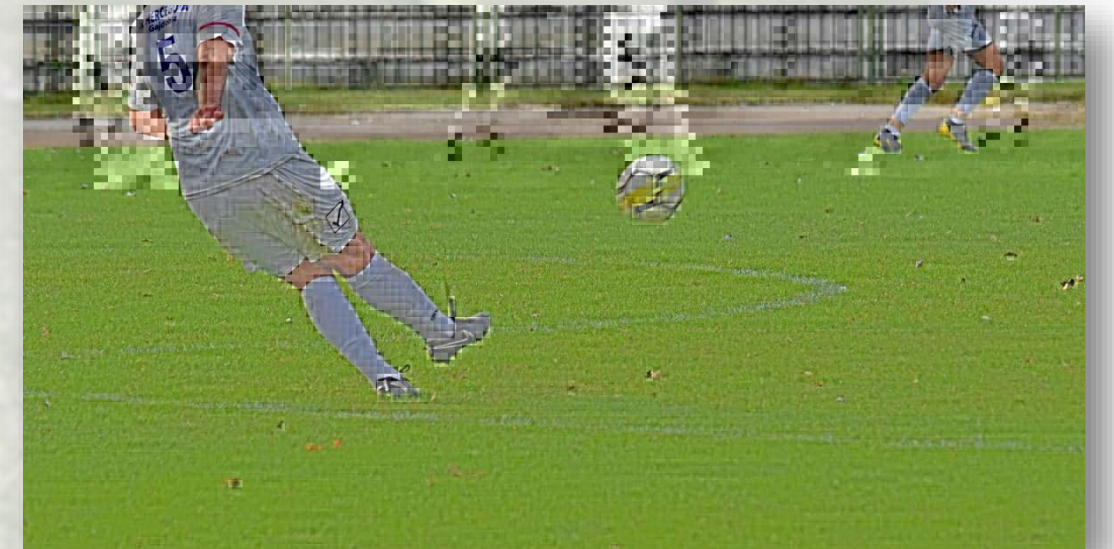Most contributing part
of 2D league

Introduction and Background

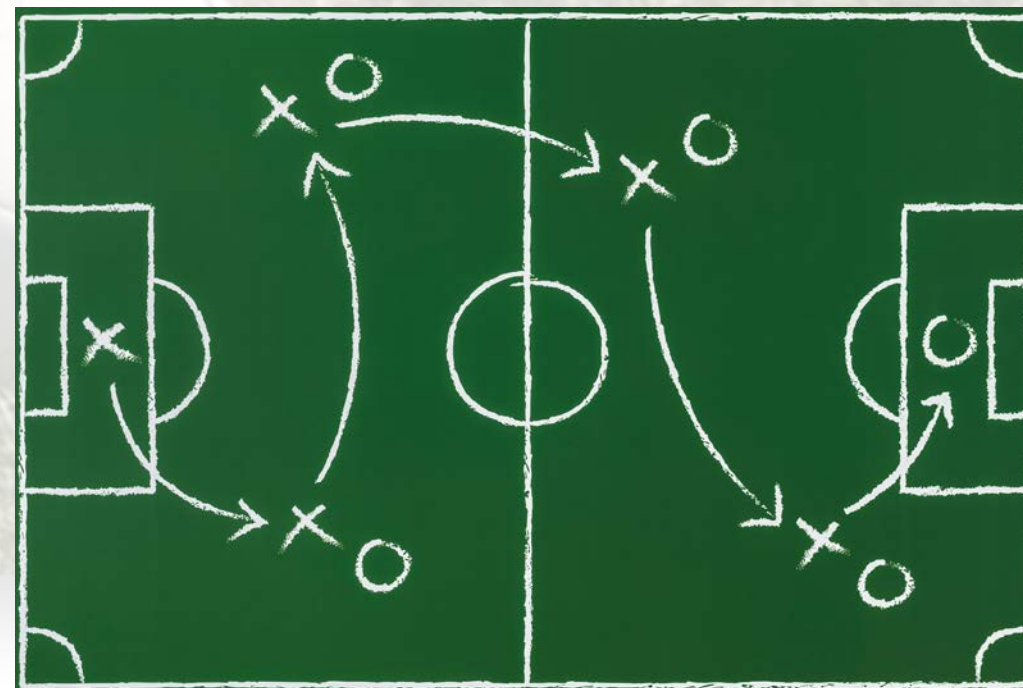Development of decision making
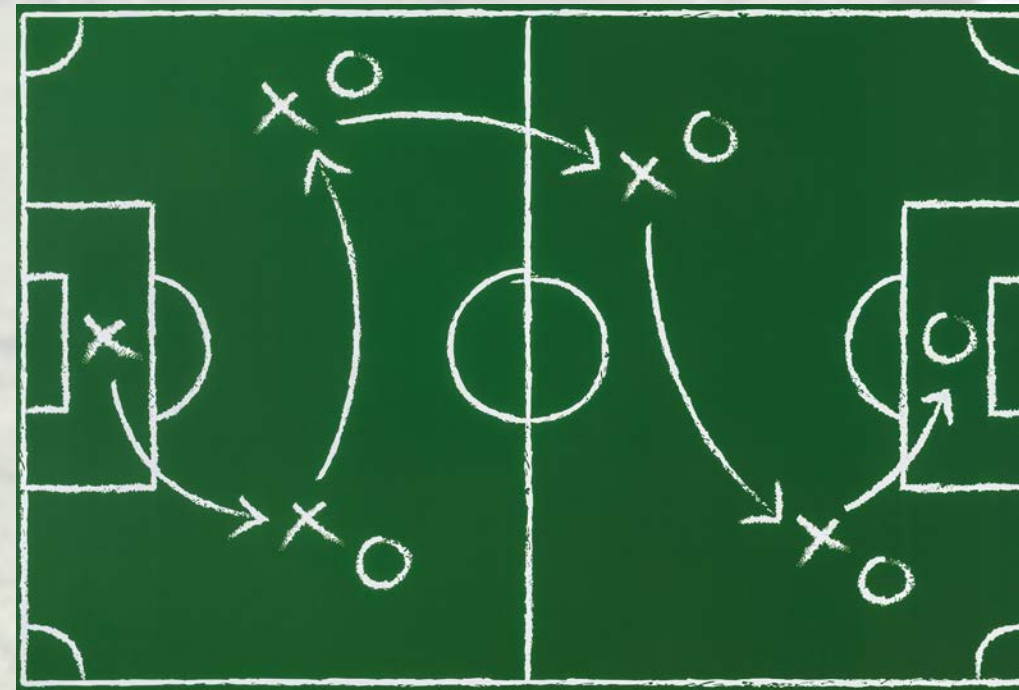
- Tactics

Dribble

Pass

- Strategies

Team play

Introduction and Background

Strategies

Team play

Analyze/Develop team's strategies by AI

Machine learning, Statistics, Simulation, etc.

Strength of Soccer Simulation 2D

- High compatibility with data analysis
  - Strategy-oriented team development

We need a lot of data!

- Repeatability
  - Agents play as long as electricity lasts

Strength of Soccer Simulation 2D

- High compatibility with data analysis
  - Strategy-oriented team development

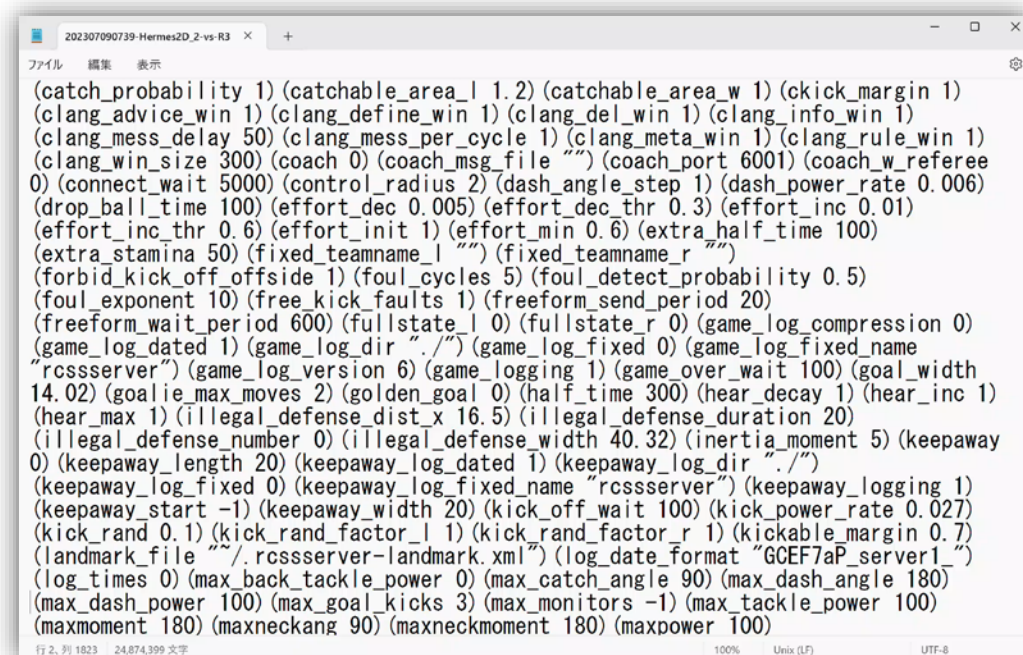- Repeatability
  - Agents play as long as electricity lasts

You can make a lot of data!
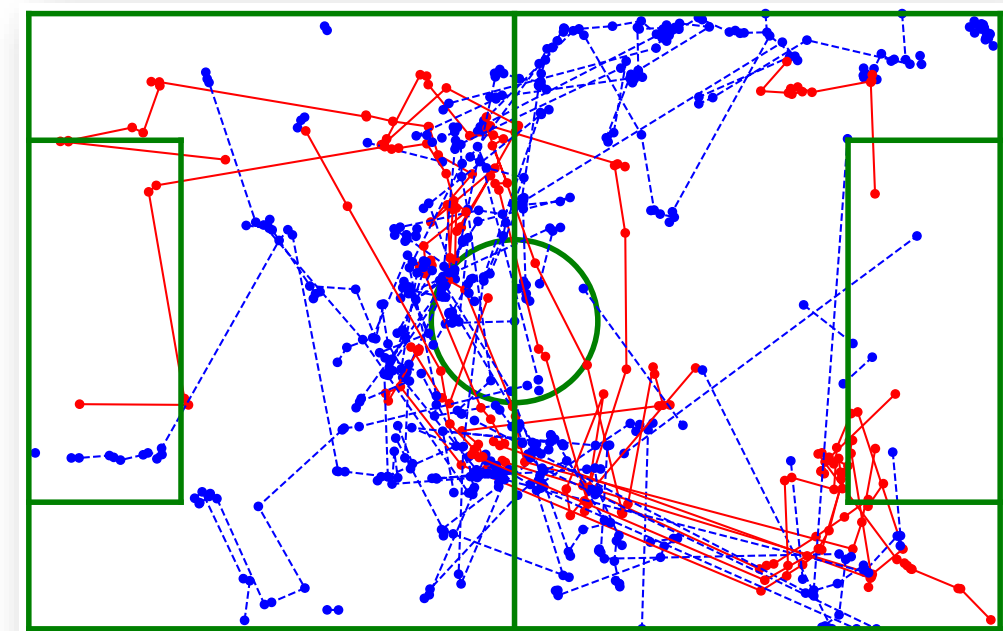
Logs Are Just Not Enough

Game logs have No event information
- Only actions are included
- Tools are necessary for extracting events such as pass, dribble, etc.

How?

Aim of Our Open-Source Development

- Instead of just providing datasets, we provide tools for generating datasets

- Two tools for: data generation and data analysis

- Provide datasets of RoboCup soccer simulation 2D

Aim of Our Open-Source Development

- Instead of just providing datasets, we provide tools for generating datasets

- Two tools for: data generation and data analysis

- Provide datasets of RoboCup soccer simulation 2D

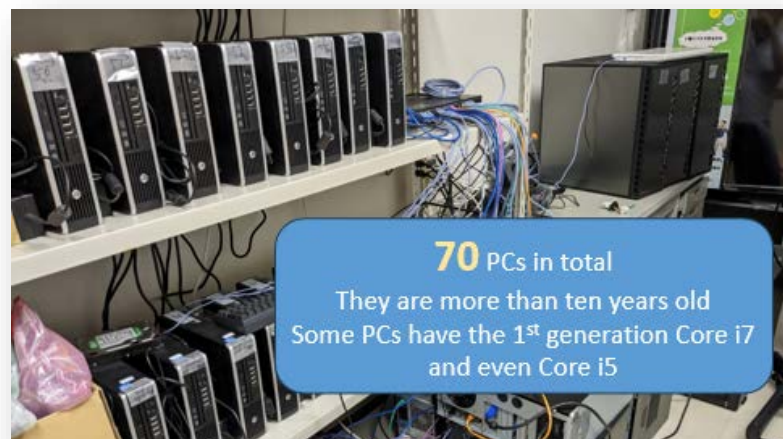# Aim of Our Open-Source Development

- Instead of just providing datasets, we provide tools for generating datasets

- Two tools for: data generation and data analysis

- Provide datasets of RoboCup soccer simulation 2D

Developed Tools

rcgamestats

・Distributed execution of soccer games

・Match scheduling
  ・Any teams in team pool with any number of games

・Result aggregation: Results database

・Summary visualization

70 PCs in total
They are more than ten years old
Some PCs have the 1st generation Core i7
and even Core i5

Confidence Interval with Goal Difference-Based Graph

left team

Developed Tools

rcg2data

- Game log conversion to CSV
- Event extraction (passes, interceptions, shots)

Type	Side1	Unum1	Time1	X1	Y1	Side2	Unum2	Time2	X2	Y2	Success
Pass	left	10	105	11.0433	32.3346	left	5	113	-2.7	30.9469	true
Pass	left	5	125	4.4467	29.5977	left	10	128	9.2386	31.489	true
Pass	left	10	143	16.6419	31.1778	left	5	152	0.0918	26.3957	true
Pass	left	5	161	4.2961	25.8656	left	6	169	2.7422	12.9022	true
Interception	left	6	173	4.7946	15.0694	right	7	174	6.834	15.8583	true
Pass	right	7	176	6.5191	16.1724	right	9	182	3.9038	24.3288	true
Pass	right	9	183	3.6503	24.2566	right	7	187	3.1349	19.623	true
Pass	right	7	188	3.1646	19.682	right	9	191	3.0246	25.26	true
Pass	right	9	204	-7.1198	24.3199	right	11	208	-10.6632	17.1693	true
Pass	right	11	209	-10.9034	17.1474	right	9	212	-8.9923	22.7204	true
Pass	right	9	212	-8.9923	22.7204	right	11	216	-12.6582	18.0909	true
Pass	right	11	227	-16.2424	16.4082	right	6	234	-15.0141	2.6168	true
Pass	right	6	234	-15.0141	2.6168	right	8	243	-13.1596	-17.0355	true
Pass	left	7	297	-24.0591	-22.1905	left	6	301	-19.5139	-14.1352	true
Pass	left	6	310	-14.5264	-14.1491	left	7	312	-16.7063	-18.5533	true

Provided Dataset

- **13 teams** from RoboCup 2021
  - Alice2021, CYRUS, FRA-UNIted, HELIOS2021, HfutEngine2021, ITAndroids, Jyo_sen2021, MT2021, Oxsy, Persepolis, RoboCin, ThunderLeague, and YuShan2021
- **1,000 round-robins**
- **78,000 matches** in total
- About 780,000 minutes in total
  - Equivalent to **8,667 human soccer matches**
- 500GB data size (zip-compressed)
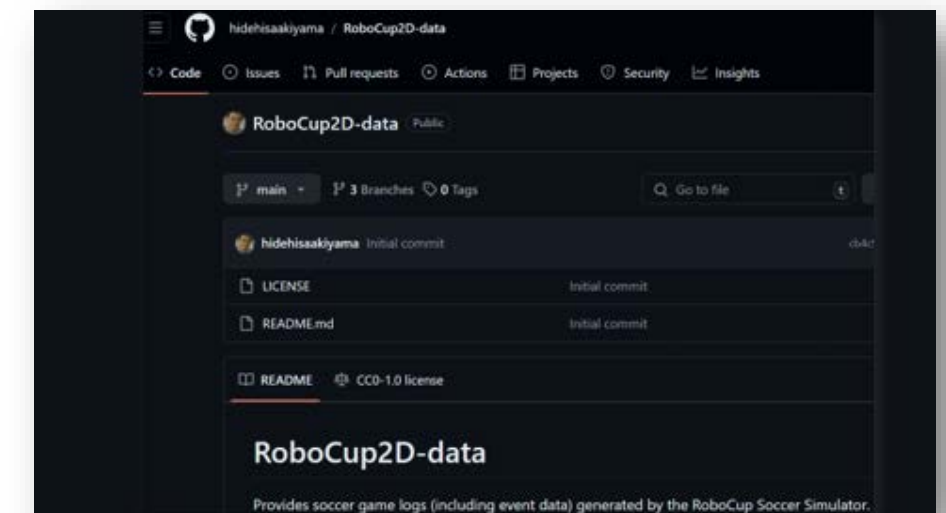
Saved Data (on a temporary server)

Index of /robocupdata/rc2021-roundrobin

Name	Last modified	Size	Description
Parent Directory		-	
anonymous/	2021-12-06 10:33	-	
check-cycle.log	2021-12-06 03:02	32K	
check-cycle.log.tar	2021-12-06 18:59	40K	
check-cycle.sh	2021-12-05 19:47	221	
count-files.sh	2021-12-06 20:28	159	
mkdirs.sh	2021-12-02 12:33	227	
normal-archived/	2023-12-22 17:22	-	
normal-csv/	2023-12-25 17:56	-	
normal/	2022-11-22 15:22	-	
teams	2021-12-02 12:33	151	

Apache/2.4.52 (Ubuntu) Server at alab.ise.ous.ac.jp Port 80

Saved Data (on a temporary server)

Index of /robocupdata/rc2021-roundrobin/normal-archived

Name	Last modified	Size	Description
Parent Directory		-	
alice2021-aras2021.tar	2021-12-06 17:31	2.9G	
alice2021-cyrus2021.tar	2021-12-06 17:34	2.9G	
alice2021-fraunited2021.tar	2021-12-06 17:38	3.2G	
alice2021-helios2021.tar	2021-12-06 17:42	2.7G	
alice2021-hfutengine2021.tar	2021-12-06 17:45	2.7G	
alice2021-itandroids2021.tar	2021-12-06 17:49	2.8G	
alice2021-jyo_sen2021.tar	2021-12-06 17:52	2.8G	
alice2021-mt2021.tar	2021-12-06 17:56	2.8G	
alice2021-oxsy2021.tar	2021-12-06 18:02	4.3G	
alice2021-persepolis2021.tar	2021-12-06 18:05	2.8G	
alice2021-robocin2021.tar	2021-12-06 18:08	2.8G	
alice2021-yushan2021.tar	2021-12-06 18:12	2.7G	
aras2021-alice2021.tar	2021-12-06 18:16	2.9G	
aras2021-cyrus2021.tar	2021-12-06 18:20	3.0G	
aras2021-fraunited2021.tar	2021-12-06 18:24	3.3G	
aras2021-helios2021.tar	2021-12-06 18:27	2.8G	
aras2021-hfutengine2021.tar	2021-12-06 18:31	2.8G	
aras2021-itandroids2021.tar	2021-12-06 18:35	2.9G	
aras2021-jyo_sen2021.tar	2021-12-06 18:38	2.9G	

Saver)

| File	Date	Size
persepolis2021-helios2021.tar	2021-12-07 02:08	2.8G
persepolis2021-hfutengine2021.tar	2021-12-07 02:12	2.8G
persepolis2021-itandroids2021.tar	2021-12-07 02:15	2.9G
persepolis2021-fyo_sen2021.tar	2021-12-07 02:19	2.9G
persepolis2021-mt2021.tar	2021-12-07 02:23	2.9G
persepolis2021-oxsy2021.tar	2021-12-07 02:29	4.5G
persepolis2021-robocin2021.tar	2021-12-07 02:33	2.9G
persepolis2021-yushan2021.tar	2021-12-07 02:36	2.8G
robocin2021-alice2021.tar	2021-12-07 02:40	2.8G
robocin2021-aras2021.tar	2021-12-07 02:44	2.9G
robocin2021-cyrus2021.tar	2021-12-07 02:48	2.9G
robocin2021-fraunited2021.tar	2021-12-07 02:52	3.2G
robocin2021-helios2021.tar	2021-12-07 02:55	2.7G
robocin2021-hfutengine2021.tar	2021-12-07 02:59	2.7G
robocin2021-itandroids2021.tar	2021-12-07 03:03	2.8G
robocin2021-fyo_sen2021.tar	2021-12-07 03:06	2.8G
robocin2021-mt2021.tar	2021-12-07 03:10	2.8G
robocin2021-oxsy2021.tar	2021-12-07 03:16	4.4G
robocin2021-persepolis2021.tar	2021-12-07 03:20	2.9G
robocin2021-yushan2021.tar	2021-12-07 03:23	2.7G
yushan2021-alice2021.tar	2021-12-07 03:27	2.7G
yushan2021-aras2021.tar	2021-12-07 03:31	2.8G
yushan2021-cyrus2021.tar	2021-12-07 03:35	2.9G
yushan2021-fraunited2021.tar	2021-12-07 03:39	3.3G
yushan2021-helios2021.tar	2021-12-07 03:43	2.7G
yushan2021-hfutengine2021.tar	2021-12-07 03:47	2.7G
yushan2021-itandroids2021.tar	2021-12-07 03:50	2.8G
yushan2021-fyo_sen2021.tar	2021-12-07 03:54	2.8G
yushan2021-mt2021.tar	2021-12-07 03:57	2.8G
yushan2021-oxsy2021.tar	2021-12-07 04:03	4.3G
yushan2021-persepolis2021.tar	2021-12-07 04:07	2.8G

Case Study: Ball Interception Analysis

Example Analysis 1

Utilization of event data to estimate probability distributions of ball interceptions
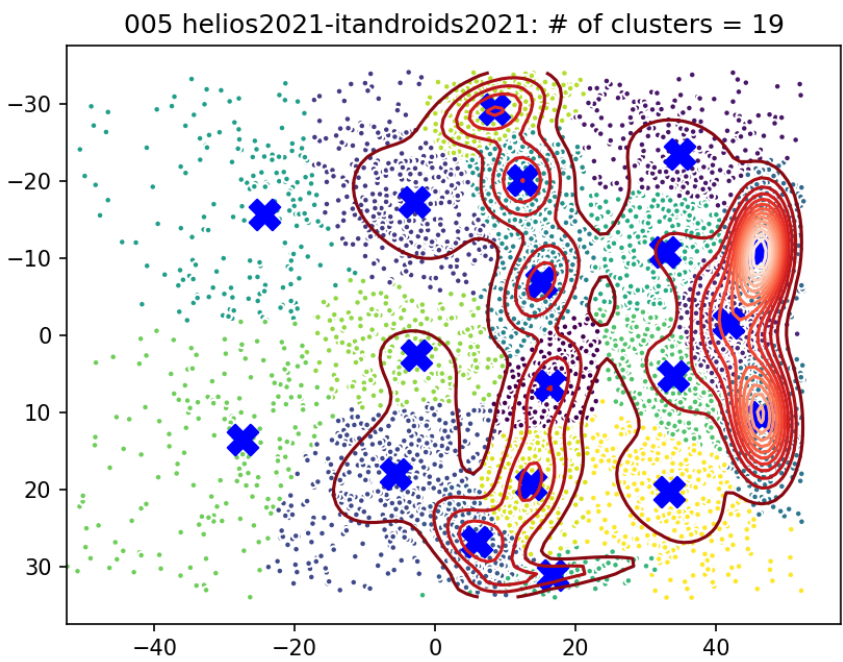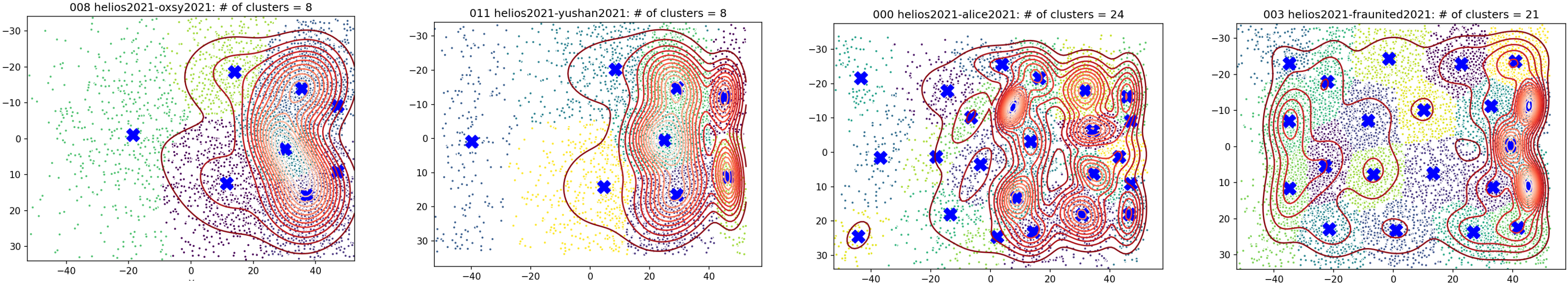
Example Analysis 2

Clustering of the estimated distribution
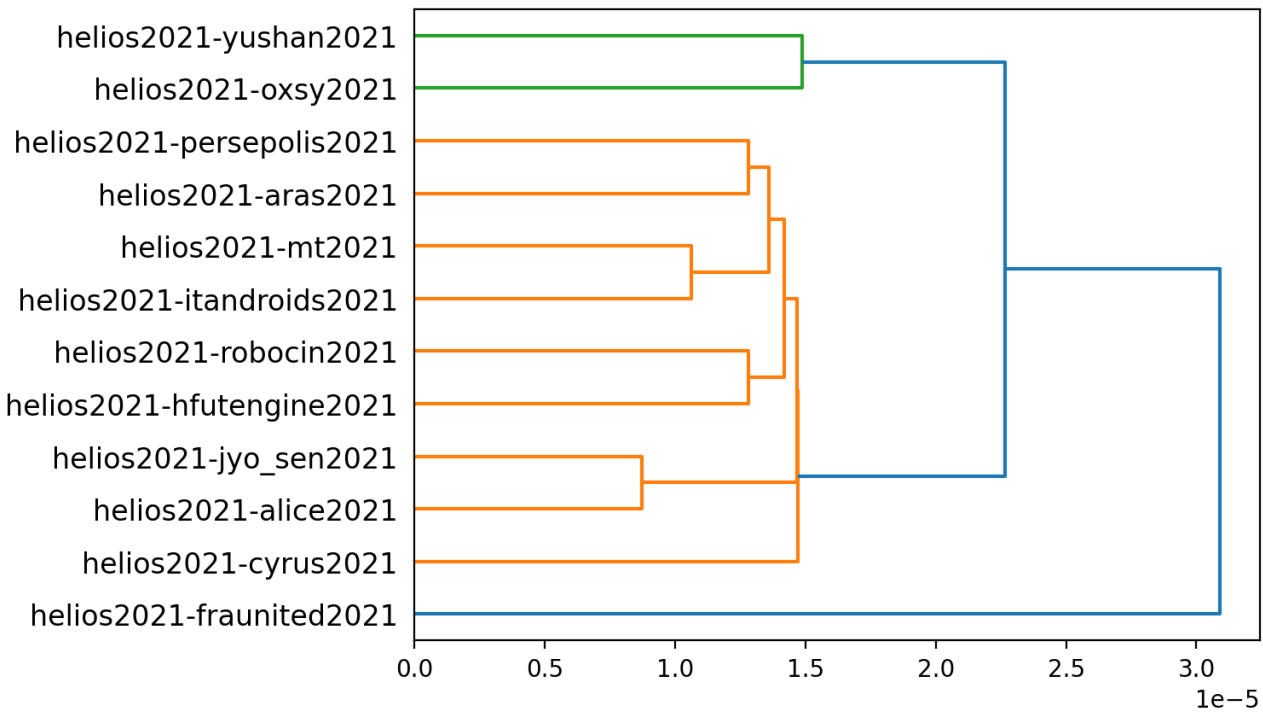
Conclusion

- Software tools were developed for data generation
- Large dataset was provided
- Analytical potential was shown as examples

Game dataset will be available after competition