DOI: 10.15439/2021F48
ISSN 2300-5963 ACSIS, Vol. 25

StarCraft strategy classification of a large human versus human game replay dataset

Štefan Krištofík, Peter Malík
Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 845 07 Bratislava, Slovakia
Email: stefan.kristofik@savba.sk

Matúš Kasáš
Tatramed Software
Líščie údolie 9, 841 01 Bratislava, Slovakia

Abstract—**Real-time strategy games are popular in AI research and education. Among them, Starcraft: Brood War (SCBW) is particularly well known. Recently, the largest known SCBW game replay dataset STARDATA was published. We classify player strategies used in the dataset for all 3 playable races and all 6 match-ups. We focus on early to mid-game strategies in matches less than 15 minutes long. By mapping the classified strategies to replay files, we label the files of the dataset and make the labeled dataset available.**

I. Introduction

In a competitive one on one real-time strategy (RTS) game environment, players try to outsmart and defeat their opponent by using superior strategy. This involves planning and execution of various tasks like good army composition, military unit placement, effective combat, scouting, territory expansion. RTS are considered very challenging for AI. The main reasons are partial observability of the game state (visible only near own units and structures) and huge complexity resulting from an overwhelming number of possible player actions [2].

StarCraft: Brood War (SCBW) released in 1998 (Fig. 1) is the most successful [9] and widely known RTS game in both the competitive and AI research communities. The competitive side has been praising the unique and fair balance [8] of all 3 playable races (Protoss, Terran, Zerg) which is rarely accomplished in gaming and is one of the main reasons for the game's longevity. The research side benefits from this longevity and over the years many useful tools were developed to help with AI research. Current goal of ongoing research is the continuous improvement of agents and ultimately overcoming human expert players on a consistent basis, which they are not yet capable of as evidenced by the results of recent AI game competitions [3], [4], [9].

One of possible approaches to agent improvement is machine learning from the past matches. SCBW allows archiving of played matches in the form of replay files. Over the years, a vast amount of such data was accumulated. However, it is scattered among many sources with various levels of quality. In case of SCBW, for machine learning purposes a dataset of game replay files should meet multiple requirements to be considered viable, mainly diversity, universality and validity [1], [8]. Recently, a high quality SCBW replay dataset called STARDATA which adheres to these requirements was published by Facebook [1]. It is a collection of 65646 game replays and is the largest dataset compiled to date. Other smaller datasets were introduced previously [5], [6], [7].

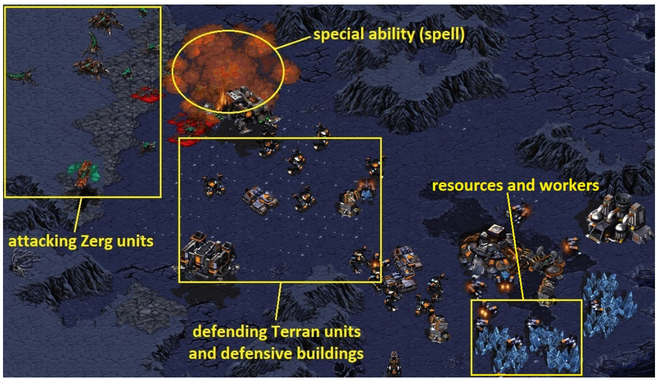


Fig. 1. StarCraft: Brood War

Following are the main contributions of this work. We classify strategies used by both players in each match from STARDATA and label the replay files with identified strategies. We make the labeled dataset publicly available [1].

II. Strategy classification

This work deals with strategy classification from the SCBW replay dataset STARDATA. Strategy classification was identified as one of the tasks which the dataset is suitable for [1]. Having such information could be helpful for future machine learning attempts and improvement of SCBW AI agents.

Some attempts at strategy classification from SCBW replays were conducted [5], [6], [7], [1], [8], but the results of each work were quite limited or very specific in some way (dataset size, quality, number of strategies, races).

In this work, we expand the idea of [8] where strategies were extracted from STARDATA, but only for the Terran race. We now consider and classify strategies of all 3 playable races in all 6 match-ups. We provide strategy information in the form of labeled files. The classification and labeling process is shown in Fig. 2 and described next.

This work is supported by Slovak national project VEGA 2/0155/19.

[1] https://drive.google.com/file/d/1mpQZoNN51iv7UX-IBanRsz0756AKO9Y3

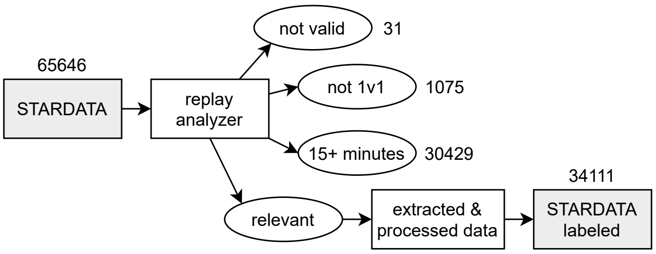


Fig. 2. Strategy classification and labeling. Numbers=amounts of replay files

TABLE I
SELECTED STRATEGY DEFINING STRUCTURES AND UNITS

Race	Structures	Units
Terran	Academy, Armory, Comm. Center, Com. Station, Con. Tower, Eng. Bay, Factory, Machine Shop, Science Fac., Starport, Refinery	Marine, Medic, Firebat, Vulture, Goliath, Siege Tank, Wraith
Protoss	Nexus, Gateway, Forge, Stargate, Cyber. Core, Templar Arch., Stargate, Robotics Facility	Zealot, Dragoon, High Tem., Dark Tem., Carrier
Zerg	Hatchery, Lair, Spire, Spawning Pool, Hydralisk Den	Zergling, Lurker, Scourge Hydralisk, Mutalisk

A. Dataset cleaning and filtering

In this work, we are interested in valid competitive 1v1 matches and early to mid-game strategies. For the match length threshold, we chose 15 minutes for the same reasons as described in [8].

We use our own software tool called BWAPI replay analyzer to automatically process the original replay files of STARDATA. It utilizes BWAPI [2], an open source API for SCBW to interact with the game engine and play back replay files, one by one, gathering useful information from each match.

1) Validation: First, each replay is checked for validity; if it can run correctly in BWAPI. Invalid replay files are removed from the relevant file pool and excluded from further processing, thus cleaning the dataset of them.

2) Filtering: Next, if a replay is valid it is checked for competitiveness and length. Only matches less than 15 minutes long and including 2 players are kept in the relevant file pool.

Dataset cleaning and filtering results are visible in Fig. 2. 34111 replays (52 % of the original dataset) remained in the pool of replays relevant for strategy classification.

B. Raw information extraction

To extract raw detailed information from relevant replay files and store in into json files (one json for each replay file), we created a modified version of the replay extractor for the Terran race introduced in [8]. The tool can now extract information from all 6 match-ups, automatically categorize it by match-up and map it to replay files.

C. Information processing

We further process the raw information and prepare it for strategy classification. SCBW strategies can be characterized mainly by [8] a) build orders - sequences of structure construction, and b) army compositions - lists of backbone unit types, i.e., the most used unit types. To collect the information about a) and b), we extract the following from each json file:

- Basic info: file name, player names and races, match length, map name, winner [8].
- For both players: count of all structure types.
- For both players: relative count of all unit types per minute. This is computed as total count divided by match length in minutes.
- For each structure and unit type: timestamp of the first occurrence.

The information is consolidated into 6 csv files, one for each match-up, each storing one match per line.

D. Strategy classification

We classify strategies of all 3 races in all 6 match-ups. Because each race would use different set of strategies against different opponents, we divide strategies into 9 categories: PP, PT, PZ (for Protoss), TT, TP, TZ (for Terran), ZZ, ZP, ZT (for Zerg). For example, Protoss would use different set of strategies against Terran (PT) than against Zerg (PZ).

Based on our domain knowledge and experience with the game, we have selected a set of most important structures and units from each race's repertoire which will be used to define various strategies. The list is shown in Table I.

Let s be the number of selected structure types in Table I. Considering only structures listed in Table I, we compute for each match in each csv file the build order as follows:

- Assign value 1 to the first structure type to be constructed by a player during a match.
- Assign value 2 to the second, 3 to third, etc., up to s.
- Assign value $s+1$ to all structure types never constructed during a match.

Let u be the number of selected unit types in Table I. Considering only units listed in Table I, we compute for each match in each csv file the unit frequency statistics as follows:

- Assign values from 1 up to u depending on the relative frequency during a match. Unit types created with higher frequency get lower values and vice versa.
- Assign value $u+1$ to all unit types never created during a match.

For strategy classification, we treat STARDATA as unlabeled data because strategies used by both opponents are unknown. We perform classification by the K-Means clustering algorithm. The goal is to identify regularities in the data. By grouping similar data into clusters, we can differentiate between various strategies. Each cluster will represent a distinct strategy. After careful adjusting and result inspection, we chose 10 as the target number of clusters for the algorithm per strategy category. This guaranteed sufficient diversity of clusters and also sufficient abundance of replays per cluster. The algorithm produces differently sized clusters. The more

[2] https://github.com/bwapi/bwapi

popular a strategy is the larger the cluster representing it will be. The outcome is a total of 90 identified strategies, 30 for each race, 10 per category.

III. Results

A small sample of classified strategies is shown in Fig. 3. Complete results with all 90 strategies are available [3]. Strategy distributions are shown in Fig. 4.

A. Strategy descriptions

Strategies in Fig. 3 are named based on categories listed in II-D and the cluster number assigned by the K-Means algorithm. For example, TP6 is a Terran player strategy used against Protoss with the cluster number 6 assigned by K-Means. The amounts of players that used each strategy are shown in column *count*. Columns *average structure order* show the build order values (explained in II-D) averaged over all the matches in each cluster. Columns *average unit frequency* show the unit frequency values (explained in II-D) averaged over all the matches in each cluster. A short verbal description of strategies is given in the last column. Descriptions focus on different aspects of strategies. In general, we are interested in the following information about each strategy:

1) Most used units: Examples: In strategy PP9, *DZ* means the most used units are Dragoons and Zealots. In ZT7, *M* means the most used unit is Mutalisk.

2) Other used units: Example: In strategy TT4, *often WMV* means Wraiths, Marines and Vultures are used very often, but are not the most used units.

3) Significant structures: If a player has built some particular structures, it may indicate they are going to produce some specific units excluded from the list in Table I. Examples: Robotics Facility built by Protoss players in strategy PP0, Lair built by Zerg players in strategy ZP6.

4) Economic expansion strategies: These try to expand economically very early and gain an income advantage over the opponent. Examples: *fast exp* indicates this type of strategy, *late exp* does not indicate it.

5) Rush strategies: These try to end the match as soon as possible by attacking or pressuring the opponent very early and catching them unprepared. Examples: *Cannon rush* in strategy PZ3, *Z rush* in ZP1.

B. Discussion

The results for each of 3 races show good variety among identified strategies. Not only 'normal' strategies are represented, but also some rush as well as economic strategies are in the mix. However, it also indicates that some tweaking of strategy defining features might be needed. For example, Carriers for Protoss are almost never used in any strategy, but Spawning Pools for Zerg are almost always built.

The results in Fig. 4 prove the variety of identified strategies for all 3 races is good which was achieved by correct selection of clustering parameters. For each race and each match-up, one can clearly identify few favorite strategies (e.g., PP0, TZ7, ZZ4) as well as those less popular (e.g., PT6, TZ0, ZZ8).

IV. Dataset labeling

We label the original STARDATA replay files by adding the following information:

- Strategies for both players.
- Match-up (can be inferred from the strategies).
- Winner flag.

Example original file: *bwrep_0xi84.rep*.

Labeled file: *bwrep_0xi84_TZ7_ZT2W.rep*.

The original unique replay ID number is preserved. Player 1 was Terran and used strategy TZ7. Player 2 was Zerg and used strategy ZT2. The winner was Player 2, indicated by the symbol *W* after their strategy.

V. Conclusion

We classify player strategies used in StarCraft: Brood War replay files from the largest known unlabeled dataset called STARDATA and label the replay files. The replay files in the labeled version now offer information about match-up and strategies used by both players and also identify the winning player. While original STARDATA may be used for unsupervised learning, in machine learning, it is always beneficial to have more options. Having the above information available makes the labeled version useful for supervised learning. We make the labeled dataset available for future machine learning attempts for StarCraft AI agent training and improvement.

References

[1] Z. Lin, J. Gehring, V. Khalidov and G. Synnaeve, "STARDATA: A StarCraft AI Research Dataset," *13th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2017,* pp. 50–56, arXiv:1708.02139.

[2] S. Ontañon, G. Synnaeve, A. Uriarte, F. Richoux, D. Churchill and M. Preuss, "A Survey of Real-Time Strategy Game AI Research and Competition in StarCraft," *IEEE Transactions on Computational Intelligence and AI in games, IEEE Computational Intelligence Society, 2013,* 5(4), pp. 1–19, doi: 10.1109/TCIAIG.2013.2286295.

[3] Mi. Čertický, D. Churchill, K.-J. Kim, Ma. Čertický and R. Kelly, "StarCraft AI Competitions, Bots and Tournament Manager Software," *IEEE Transaction on Games, 2018,* 11(3), pp. 227–237, doi: 10.1109/TG.2018.2883499.

[4] O. Vinyals, I. Babuschkin et al., "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature, 2019,* 575, pp. 350–354, doi: 10.1038/s41586-019-1724-z.

[5] B. G. Weber and M. Mateas, "A data mining approach to strategy prediction," *IEEE Symposium on Computational Intelligence and Games, 2009,* pp. 140-147, doi: 10.1109/CIG.2009.5286483.

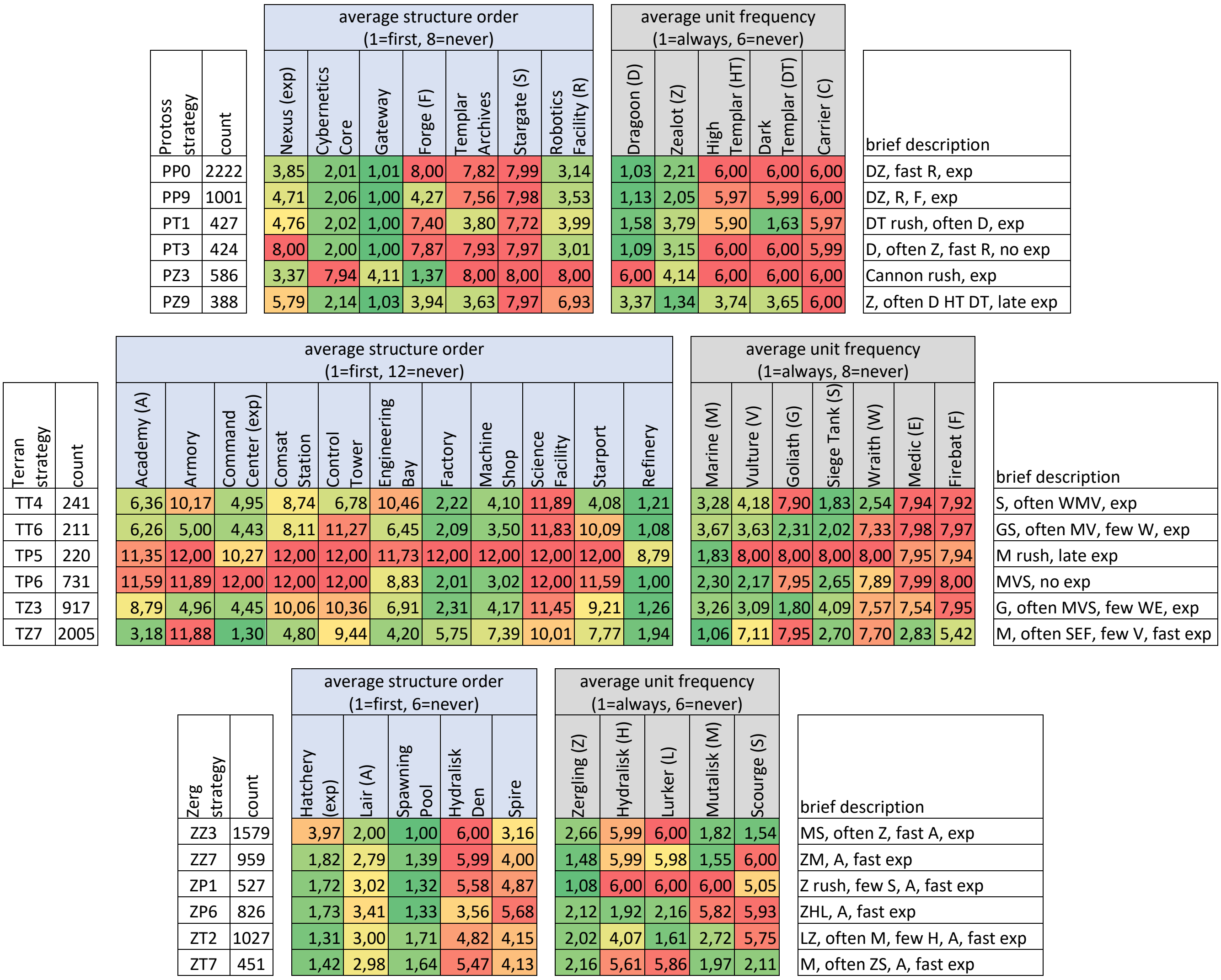
[6] H. C. Cho, K. J. Kim and S. B. Cho, "Replay-based strategy prediction and build order adaptation for StarCraft AI bots," *IEEE Conference on Computational Intelligence in Games (CIG), 2013,* pp. 1-7, doi: 10.1109/CIG.2013.6633666.

[7] G. Synnaeve and P. Bessière, "A Dataset for StarCraft AI & an Example of Armies Clustering," *Artificial Intelligence in Adversarial Real-Time Games, 2012,* arXiv:1211.4552.

[8] Š. Krištofík, P. Malík, M. Kasáš, Š. Neupauer, "StarCraft agent strategic training on a large human versus human game replay dataset," *Federated Conference on Computer Science and Information Systems, FedCSIS 2020,* 21, ACSIS, pp. 391–399, doi: 10.15439/2020F178.

[9] M. Świechowski, "Game AI Competitions: Motivation for the Imitation Game-Playing Competition," *Federated Conference on Computer Science and Information Systems, FedCSIS 2020,* 21, ACSIS, pp. 155–160, doi: 10.15439/2020F126.

[3] https://drive.google.com/file/d/184ZV5VCzj75avTxjxSM1qLncRjmzs_9o



		average structure order (1=first, 8=never)							average unit frequency (1=always, 6=never)					
Protoss strategy	count	Nexus (exp)	Cybernetics Core	Gateway	Forge (F)	Templar Archives	Stargate (S)	Robotics Facility (R)	Dragoon (D)	Zealot (Z)	High Templar (HT)	Dark Templar (DT)	Carrier (C)	brief description
PP0	2222	3,85	2,01	1,01	8,00	7,82	7,99	3,14	1,03	2,21	6,00	6,00	6,00	DZ, fast R, exp
PP9	1001	4,71	2,06	1,00	4,27	7,56	7,98	3,53	1,13	2,05	5,97	5,99	6,00	DZ, R, F, exp
PT1	427	4,76	2,02	1,00	7,40	3,80	7,72	3,99	1,58	3,79	5,90	1,63	5,97	DT rush, often D, exp
PT3	424	8,00	2,00	1,00	7,87	7,93	7,97	3,01	1,09	3,15	6,00	6,00	5,99	D, often Z, fast R, no exp
PZ3	586	3,37	7,94	4,11	1,37	8,00	8,00	8,00	6,00	4,14	6,00	6,00	6,00	Cannon rush, exp
PZ9	388	5,79	2,14	1,03	3,94	3,63	7,97	6,93	3,37	1,34	3,74	3,65	6,00	Z, often D HT DT, late exp

		average structure order (1=first, 12=never)											average unit frequency (1=always, 8=never)							
Terran strategy	count	Academy (A)	Armory	Command Center (exp)	Comsat Station	Control Tower	Engineering Bay	Factory	Machine Shop	Science Facility	Starport	Refinery	Marine (M)	Vulture (V)	Goliath (G)	Siege Tank (S)	Wraith (W)	Medic (E)	Firebat (F)	brief description
TT4	241	6,36	10,17	4,95	8,74	6,78	10,46	2,22	4,10	11,89	4,08	1,21	3,28	4,18	7,90	1,83	2,54	7,94	7,92	S, often WMV, exp
TT6	211	6,26	5,00	4,43	8,11	11,27	6,45	2,09	3,50	11,83	10,09	1,08	3,67	3,63	2,31	2,02	7,33	7,98	7,97	GS, often MV, few W, exp
TP5	220	11,35	12,00	10,27	12,00	12,00	11,73	12,00	12,00	12,00	12,00	8,79	1,83	8,00	8,00	8,00	8,00	7,95	7,94	M rush, late exp
TP6	731	11,59	11,89	12,00	12,00	12,00	8,83	2,01	3,02	12,00	11,59	1,00	2,30	2,17	7,95	2,65	7,89	7,99	8,00	MVS, no exp
TZ3	917	8,79	4,96	4,45	10,06	10,36	6,91	2,31	4,17	11,45	9,21	1,26	3,26	3,09	1,80	4,09	7,57	7,54	7,95	G, often MVS, few WE, exp
TZ7	2005	3,18	11,88	1,30	4,80	9,44	4,20	5,75	7,39	10,01	7,77	1,94	1,06	7,11	7,95	2,70	7,70	2,83	5,42	M, often SEF, few V, fast exp

		average structure order (1=first, 6=never)					average unit frequency (1=always, 6=never)					
Zerg strategy	count	Hatchery (exp)	Lair (A)	Spawning Pool	Hydralisk Den	Spire	Zergling (Z)	Hydralisk (H)	Lurker (L)	Mutalisk (M)	Scourge (S)	brief description
ZZ3	1579	3,97	2,00	1,00	6,00	3,16	2,66	5,99	6,00	1,82	1,54	MS, often Z, fast A, exp
ZZ7	959	1,82	2,79	1,39	5,99	4,00	1,48	5,99	5,98	1,55	6,00	ZM, A, fast exp
ZP1	527	1,72	3,02	1,32	5,58	4,87	1,08	6,00	6,00	6,00	5,05	Z rush, few S, A, fast exp
ZP6	826	1,73	3,41	1,33	3,56	5,68	2,12	1,92	2,16	5,82	5,93	ZHL, A, fast exp
ZT2	1027	1,31	3,00	1,71	4,82	4,15	2,02	4,07	1,61	2,72	5,75	LZ, often M, few H, A, fast exp
ZT7	451	1,42	2,98	1,64	5,47	4,13	2,16	5,61	5,86	1,97	2,11	M, often ZS, A, fast exp

Fig. 3. A small sample of Protoss (top), Terran (middle) and Zerg (bottom) strategies

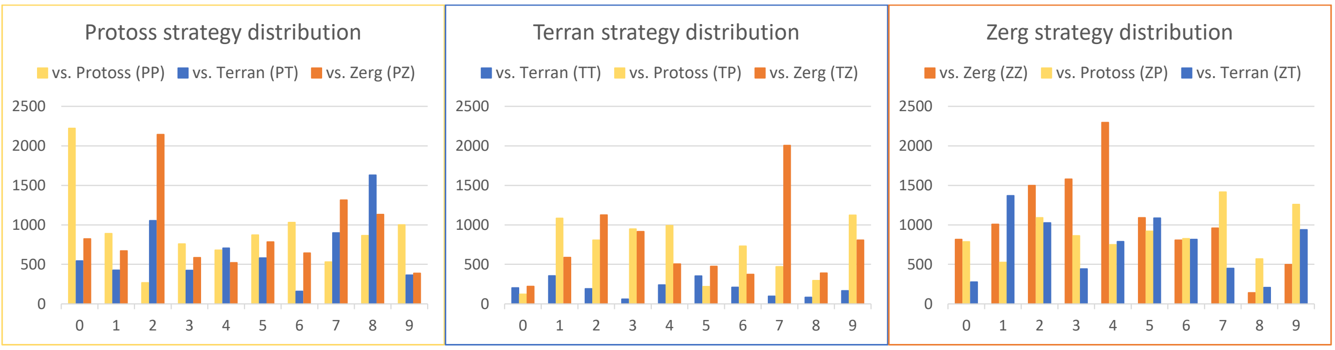


Fig. 4. Strategy distribution