Starcraft 2 Performance: An In-Depth Look At In-Game Telemetry And Player Rank

Authors: Chuyun Kuang, Minsuh (Eric) Lim, Shawn Xu, Linda Ding
Georgetown University

Abstract

In the burgeoning field of esports, the intersection of machine learning and video game analysis offers new opportunities for discovering complex patterns in player performance and strategy development. Previous research from Thomas et al. has focused on complex skill learning through video games, and Google's Starcraft AI AlphaStar has shown Starcraft II as a focal point for machine learning and analytics. We will dissect the multifaceted nature of esports with a focus on Starcraft II.

Introduction & Motivations

In the rapidly evolving landscape of digital entertainment, competitive gaming, commonly known as esports, has emerged as a phenomenon captivating millions worldwide. The roots of esports trace back to the early days of video gaming, where the spirit of competition found its expression in local arcades and small tournaments. Despite its humble beginnings, the 21st century marked a pivotal moment in gaming history as internet connectivity and advancements in technology catalyzed the exponential growth of esports. Today, esports encompasses a wide range of games, genres, and formats; it has fostered a global community of professional players, enthusiasts, and spectators, reaching a level of ubiquity that led to its inauguration into the Olympics in 2023. Amidst this backdrop, analyzing esports through machine learning presents a novel approach to understanding the intricacies of competitive gaming at a granular level.

We attempt to bridge the gap between the historical evolution of esports and the analytical depth that machine learning can provide to esports. Our primary focus is on analyzing the "godfather of esports," Starcraft. Before the analysis, we will set the stage with a high-level overview of the esports scene in the 21st century. We then perform an analysis of the SkillCraft1 Master Table Dataset. Our objective is to illuminate the multifaceted nature of player performance, strategy development, and learning processes within the competitive realm of StarCraft II (SC2), thereby offering insights into the real-time strategy (RTS) genre, as well as the broader domain of esports. We aim to contribute a novel perspective to the interdisciplinary field of game studies, one that underscores the potential of data-driven research in unlocking the complexities of digital competition.

Statistical Analyses: ANOVA and Pairwise T-tests

Actions per minute (APM) is a common statistic used to describe a player's capabilities in SC2. The ability to multitask and optimize each second makes APM a reliable metric for determining a player's skill and knowledge. Additionally, screens per minute (SPM) is used to measure how quickly one moves their screen around the map. A higher SPM generally means more map information for the player, which is invaluable for RTS games where decisions are made based on incomplete information. This statistical analysis will explore pro-level matches from IEM Katowice 2016 to 2023 to examine whether certain races require different levels of APM/SPM to secure a win through ANOVA tests. The distributions of APM and SPM based on race can be found in Figures 2 and 3.

With the SkillCraft1 Dataset, we will look at how APM compared at each rank through multiple pairwise T-tests. It should be noted, however, that the SkillCraft1 Dataset contains many more variables than just APM and rank. Figure 4 shows the average APM associated with each rank. The results of the pairwise T-tests can be found in Figure 5.

Machine Learning: Regression and Classification

Next, this section will delve into the application of various machine learning models to predict two primary aspects of a player's performance: APM and player rank. The first part of the analysis employs linear regression to predict APM, leveraging its continuous nature to assess how different features influence a player's ability to execute actions rapidly and efficiently. Furthermore, we will also classify players into different ranks, a categorical outcome that encapsulates a player's overall skill. This will be done with logistic regression, support vector machines (SVMs), random forests (RFs), and XGBoost.

The correlation among the variables in the SkillCraft1 Dataset can be found in Figure 6. APM has a very strong positive correlation with SelectByHotkeys (0.815). This makes sense since selecting units or buildings would contribute to a player's APM. Secondly, APM has a strong negative correlation with ActionLatency (-0.722), which also is logically sound, since higher APM would require a player to make more actions, leading to less time between each action.

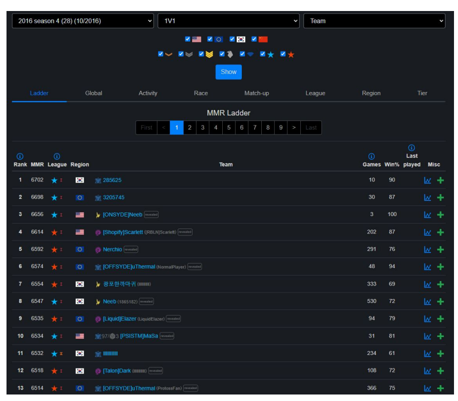

Figure 1: SC2 Ladder Tracker. Data were collected from SC2Pulse for every player in each of the four regions from season 16 (October 2016) to season 58 (April 2024).

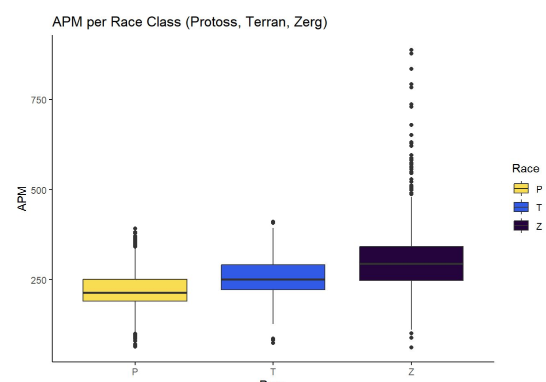

Figure 2: Boxplot of APM by Race. Zerg has the highest APM, followed by Terran, and then lastly Protoss.

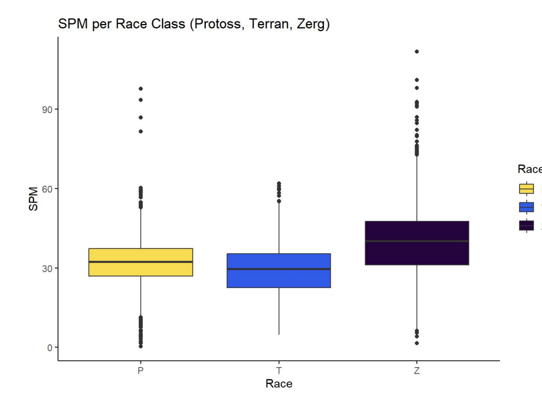

Figure 3: Boxplot of SPM by Race. Zerg has the highest SPM, followed by Protoss, and then lastly Terran.

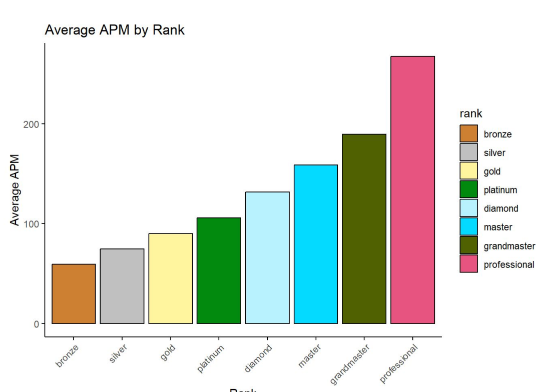

Figure 4: Histogram of Average APM by Rank.

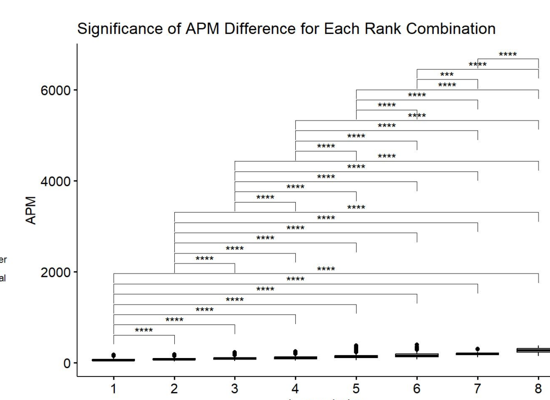

Figure 5: Visualization of pairwise T-tests of APM by Rank. The significance is represented by *** or ****, which corresponds to p-value < 0.0001 or p-value < 0.001, respectively.

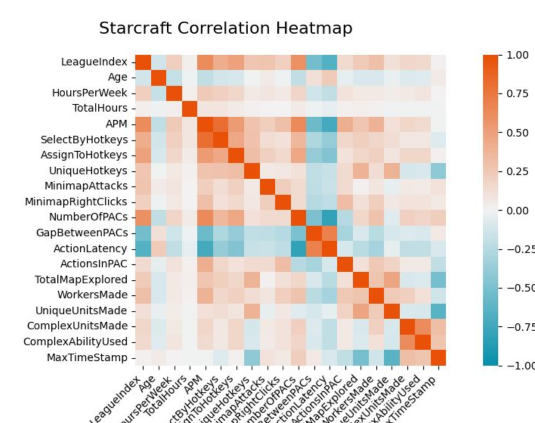

Figure 6: Correlation Matrix of Numeric Features. APM appears to have a very strong positive correlation with SelectByHotkeys and a strong negative correlation with ActionLatency.

About the Data

Based on the available filters on SC2Pulse, the fields that were saved were season, region, the player's matchmaking rating (MMR), rank, and race (Figure 1). Seasons ranged from 28 to 58, covering about eight years of data. Available regions were China, Europe, Korea, and the U.S. The player's MMR is the numeric representation of a player's skill, and is directly related to their rank. Rank refers to the different rank divisions in the ladder, which are Bronze, Silver, Gold, Platinum, Diamond, Master, and Grandmaster, ordered from worst to best. Lastly, race refers to the class a player primarily plays, with the options being Zerg, Terran, Protoss, and Random. The race Random was scraped for players who did not predominantly play a specific race. It should be noted that Random would assign one of the three other races to the player upon a match starting.

Before applying any machine learning models to our primary dataset (i.e. SkillCraft1 Master Table Dataset), we removed any rows with missing values. Upon further inspection, these rows primarily related to players classified as "Professional." However, considering that the pros are an extreme minority of the player base, this seemed to be a pretty reasonable design choice. Additionally, uninformative columns (e.g. Game ID) were dropped.

For the statistical tests, where we used the Intel Extreme Masters (IEM) Katowice data from Kaggle, we removed any game where a player chose Random as their race. This was done to ensure that the statistical tests would provide insights relating to the three playable races.

Statistical Analyses Results

Based on these ANOVA tests, keeping the skill level consistent, we found that the average APM and SPM significantly differed from each other when considering the combinations of races from this IEM dataset. This may suggest that Zerg players have to "work harder" than Protoss or Terran players.

Similar to the findings in the ANOVA tests, we can also see that APM is also an important metric for determining a player's rank, with every pairwise T-Test resulting in a statistically significant result (i.e. every paired rank's APMs had a mean not equal to zero).

Machine Learning Results

The linear regression model performed very well in predicting a player's APM, so much so that the pruned "optimal" version (with features Age, SelectByHotkeys, MinimapAttacks, MinimapRightClicks, NumberOfPACs, ActionLatency, ActionsInPAC, WorkersMade, UniqueUnitsMade, and ComplexAbilitiesUsed) performed essentially as well as the full model (R-squared=0.9755; p-value < 2.2e-16; RMSE=9.358). This serves as a testament to the well-defined and linearly separable nature of this particular aspect of SC2, which we initially saw in Figure 4.

However, our attempts to classify ranks revealed the challenges behind the nuanced and multifaceted nature of player skill. Despite applying a suite of machine learning models, none of the 24 models tested resulted in great performance in predicting a player's rank. Although the target variable was quite imbalanced, attempts at addressing this with SMOTE still proved unfruitful. The poor performance across all models highlights the complexity of capturing the essence of a player's rank through the gathered metrics, with the best model's confusion matrix shown in Figure 7. More specifically, the RF model with SMOTE balancing yielded an accuracy of 0.399 and ROC AUC Score of 0.802.

True \ Predicted	1	2	3	4	5	6	7
1	15	5	2	3	0	0	0
2	14	18	12	7	1	0	0
3	5	25	28	14	7	4	0
4	5	18	20	36	36	7	0
5	3	2	12	25	47	31	0
6	0	0	3	10	23	55	2
7	0	0	0	0	0	5	1

Figure 7: Confusion Matrix of Best Classification Model (Random Forest Model with SMOTE balancing). The model frequently misclassifies for all classes.

Conclusion

Our statistical tests showed that a player's race was an important indicator when comparing APM and SPM. Additionally, APM also seems to be an important indicator for distinguishing between the different skill levels.

The regression analysis also proved quite fruitful with our best model performing akin to the full model despite being much less complex.

On the other hand, the classification portion fell short. This may be due to the complexity of a player's rank, which may be difficult to capture with merely in-game metrics. For example, a player's playstyle (e.g. greedy vs. conservative, economical vs. aggressive) and decision making (i.e. how often does a player scout? how does the player respond to what the opponent is doing?) are difficult to represent numerically. As such, gathering a more robust dataset as well as applying deep learning may lead to better model performance.

References:
"History of Esports." CDW.Com, CDW, 10 Sept. 2021, www.cdw.com/content/cdw/en/articles/hardware/history-of-esports.html.
Larch, Author: Florian, and Florian Larch. "Emergence of Esports: Once Ridiculed, Now a Billion-Dollar Market." ISPO.Com, 24 Oct. 2023, www.ispo.com/en/sports-business/esports-history-how-it-all-began.
Thompson, Joseph J., et al. "Video game telemetry as a critical tool in the study of Complex Skill Learning." PLoS ONE, vol. 8, no. 9, 18 Sept. 2013, https://doi.org/10.1371/journal.pone.0075129.