**The Road on Becoming the Next E-Sports Professional**

By Edward Lee

The world of E-Sports is growing at a very rapid pace and many young prospects are looking towards gaming as a potential career choice. However, how does one person decide if their talent or skill level is enough to compete with the best in the world? How much time should be invested or achieved to be better than an average player? What is the threshold to decide whether one person is suitable for this career choice? What is the difference between playing the game myself and the professional players I watch on twitch.tv or other online live streaming websites?

The dataset I will be using is the SkillCraft1 Master Table Dataset Data Set from the University of California, Irvine Machine Learning Repository dated on September 20th, 2013. The dataset is based on StarCraft II replay files received from players in the lowest league to the players competing at a professional level. The replay files are from 1v1 matches running StarCraft 2 version 1.3.6.19269:291.

When I initially found the dataset, I was intrigued at how detailed every variable was recorded. I immediately set my eyes on LeagueIndex and APM because, as an avid gamer and having played StarCraft II before, your competitive rank and APM should be relatively correlated. However, before doing any exploratory data analysis, the dataset required some data wrangling and manipulation.

The dataset was fairly clean and did not require a large amount of time. The first thing I noticed was LeagueIndex == 8 had empty cells in Age, HoursPerWeek and TotalHours. The dataset was fairly large (3395 observations), so I took the mean of the other ages and inputted the mean into LeagueIndex == 8. As for HoursPerWeek and TotalHours, I took the game release date subtracted from the date the dataset was released, estimated that professional players play roughly 8 hours per day (normal working hours for any white-collar employee), multiplied by 20 days per month and 12 months per year and inputted those values accordingly. HoursPerWeek was manipulated the same way, except changed to a weekly value. After checking summary(inputData), I noticed that the maximum TotalHours had a value of 1,000,000 (which is impossible) and the mean was fairly skewed compared to the median. I checked table(inputData$TotalHours) and noticed there were many values that were impossible because they exceeded the amount of hours of when the game was released. It seemed like an input error because all the numbers, if divided by 10, became an acceptable value. For the value at 1,000,000, I divided the value by 1,000. Next, I did not want to use timestamps as my time value, so I had to manipulate the data to show minutes. The data recorded in the dataset was recorded while on fast, which is 1.5x speed the normal speed. The dataset provider listed that 1 real time second equates to 88.5 timestamps. I switched the timestamps into minutes for easier calculation and comparison for the future. Lastly, I changed all the LeagueIndex numbers into their actual rank names.

After data wrangling and manipulation, I looked further into APM and PAC. As I originally stated, my knowledge was that LeagueIndex and APM is positively correlated and the biggest factor to determine the skill level between players. However, I was not entirely correct. PAC is actually a much better skill determinant than APM. First, we need to identify the definition and difference between APM and PAC.

APM is simply 1 action of every click of the mouse onto an object in the game that produces an immediate result or the hitting a key on the keyboard to switch frames, objects (units, structures or spells), to build or upgrade a structure or to produce a unit. In theory, when a player’s APM is much higher than their opponents, the player with the higher APM should win because they would be able to gather more resources, produce more structures and units and command their army on their will. Through the dataset, the highest achieved sustained APM through an entire game is 389.80 with a median of 108 and mean of 117. The difference between an average player and, most likely a professional player, is a huge difference. However, just slightly glancing through the LeagueIndex of 8 (professional players), there are also APMs as low as roughly 146 and 179. Does APM really stand between a potential youngling looking into becoming the next E-Sports professional or is it just a number to scare you away?

The PAC (Perception-Action Cycle) variables, which include NumberOfPACS, GapBetweenPACS, ActionLatency and ActionsInPAC, may be the most important variables that can differ an average player to a professional player. PAC, in StarCraft II terms, is the shifting of a screen, then followed by at least one action before shifting to a new location on the screen. Many may ask, why are these PAC variables more important than APM? Well, APM has flaws. Although it is still a very good indicator for how many actions a player can manage per minute, it is too generalized. Players who are very good at macro-management, which is the constructing and producing of structures and units, may not be good at micro-management, which is moving your units, attacking with your units and using of unit abilities, and vice versa. Every key pressed or mouse click adds to 1 APM, but that doesn’t show what has actually been done. Players may have bound numbers 1 to their main structure and 2 to their barracks, press the keys over and over again, and still produce a very high APM, which is not accurate. However, PAC calculates the actions a player has done after shifting screen, which can more accurately show that a player has continuously changed their perspective on the game to, not only build their army and structure up, but to adjust to the game at real time. Players who are able to constantly, within low durations (GapBetweenPACS), perform high amounts of screen shifting (NumberOfPACS), high amounts of action per each shift (ActionsInPAC) and the milliseconds before their perform an action (ActionLantency), can arguably be concluded to be a better player and have a better understanding of the game.

Preliminary exploration of the dataset shows that the mean and median, as well as the range, increases as players progressed. However, the difference in mean and median from GrandMaster, the highest rank achievable online, and professional players has a much larger gap than from Master to GrandMaster. PAC is much more interesting because the gap is not as big as I thought. Although there is still a difference between GrandMaster and professional players, we can see that GrandMaster players do have the ability to compete with professional players. However, this requires more detailed data gathering, if the data is even out there. Also, I have an assumption that the majority of professional players will most likely be practicing through Blizzard’s online Battle.Net system, rather than against other professional players on a daily basis, and assuming that all professional players are at the highest rank possible for StarCraft II, which is GrandMaster, may skew the actual data of GrandMasters. However, this is just an assumption and without more detailed information about where the data came from, the conclusion is null and void.

Regardless of what correlation I looked at (Hours per Week vs APM, Total Unique Hotkeys vs Ranks, Number of PACs, etc…), the more a player spends their hours per week, the higher your rank will be. This makes sense because the more time you are invested in the game, the larger your experience and game knowledge expands, the more natural you are with the game, the better you will become. This is true for pretty much everything about education and experience. The more you are invested into anything, the more experienced, more natural and better you become with that skill or knowledge. One correlation that did not follow this trend was Minimap Right Clicks vs Minimap Attacks. Minimap Right Clicks is the clicking the bottom left minimap, which shows the entire map of the battle ground. Minimap Attacks is pressing keyboard “A”, while having units selected, then clicking on an area on the minimap. The units will automatically travel to that destination and attack all hostile units on their way. Surprisingly, the difference between Bronze to Professionals barely have any patterns. Rather, this seems to be more how the player prefers to play. This correlation helped me re-think that, although more time invested increases skill levels, there is still the human side of preference. The skill set of selecting and assigning more hotkeys at the pace of an APM or PAC, but people still have a preference on how they look at everything. An example that I thought of is basketball. There is the perfect form to shooting a basketball, but how many professional players actually shoot that way? How many immature players actually shoot in a better looking form than professional players, but do not make as many as professional players? Having a preference on how you like to do something is as important as making the skill part of yourself.

The final variables I looked at were the ones related to PAC. As I assumed, the number of PACs per each action and the time gap (in milliseconds) between each PAC increases as your rank increases. However, what I was really interested was APM vs PAC per minute. When I plotted the median APM vs median number of PACs per minute and the mean APM vs mean number of PACs per minute, both plots showed quite a straight line. I checked the cor.test for both median and mean and found that the median correlation was 0.9622201 and the mean correlation was 0.9589628. These 2 variables were significantly correlated. However, the reason why it is not 1.0 is because there was a significant difference between professional players and GrandMaster players. Not only do professional players have an amazingly high APM, they have a much higher PAC than GrandMaster players. From my analysis, they have played the game so much that they understand every detail of the game, everything comes naturally. The ability constantly maintain an APM at roughly 300 and performing roughly 27 PACs per minute is incredible. Professionals are really the best in the world and GrandMaster players, although talented and have invested a lot of time getting to that point, would need a huge edge to compete against the world’s best.

My final analysis was a SVM regression analysis to test whether LeagueIndex has a relationship to the other 19 variables. I partitioned the total data by 70% training set and 30% testing set. Through testing, an optimal cost range of seq(0.25, 0.35, by 0.01) with the optimal cost being 0.32 and optimal gamma range of seq(0.005, 0.015, by = 0.001) with the optimal gamma being 0.012 is used. After analysis, the optimal Train RMSE is 1.005571, optimal Test RMSE is 0.9757243 and optimal 5-fold CV RMSE is 0.926072. The model predicts quite well of slightly 1.0 deviates away. With this, we can conclude that each variable tested against the LeagueIndex, which is the ranks the players can achieve through Blizzard’s online Battle.net system, shows a good fit or correlation that higher ranks require a better set of skills and invested time.

However, there are some limitations to the dataset provided. There is no Win/Loss column, Win Rate column or Win rate against higher or lower ranked players. Also, it would be interesting to have a mean length per game for each rank to see if there are any disparities between low ranking games versus high ranking games. Also, we do not know which country of origin the professional players are from versus the players who submitted their data. The overall quality of players in South Korea are much stronger than the overall quality of players in North America and Europe. The data analysis would be much more interesting to see the difference in LeagueIndex of different countries around the world and how they differentiate from each other.

The analysis was initially not targeted to any company because I am trying to present that reaching a level to compete against the best players in the world, is not easy and actually requires dedication. The data analysis would be closer to a blog post for all younger generations considering professional gaming as a future career. Based on my analysis, I hope to enlighten them on how much minimum time should be invested to be able to compete against the best in the world and, if they are talented, what minimum skills must be achieved. Without the knowledge, the younger generation will never know the difference between their own talents versus the world professionals already in the business. However, after researching and reading that universities, such as University of California Irvine and New York University, have started to provide Professional E-Sports Gaming Degrees, it would be interesting to provide this analysis to universities who are training their students. I would suggest universities to create 3 different training plans to train their students in StarCraft II through the results of my analysis.

First would be a PAC and hotkey training plan. As mentioned before, PAC is the shifting of a screen, then followed by at least one action before shifting to a new location on the screen. Generally speaking, this is the act of multitasking. Students must be able to multitask and think quickly in real time because every second counts. There is a custom map in the StarCraft II called Multitasking Trainer. The map creates most likely scenarios in a real game by forcing you to do multiple actions on different parts of the map constantly. This is a great training map because if you fail to achieve any of the actions, you fail. There are different difficulty settings and challenge mode, such as increasing the game speed, increase actions required to make, etc… Other multitasking plans would be to train your students to write a report of a gaming video, while listening lyrics to their favorite song and constantly checking youtube updates for new videos. The training plans are endless, but those are a couple of examples.

Next would be an APM training plan. APM is positively correlated to PAC, so practicing with the above plan would also help your APM. However, I think there is a more interesting way to train APM. There are many flash games online, such as this one <http://www.onemorelevel.com/game/multitask>, which forces you to do multiple actions by constantly putting you into situations where you have to make an action. There are probably many more solutions similar to this, but I find this to be a great exercise.

Finally would be to separate your students by talent. While exploring the data, I noticed that there are a ton of outliers at every rank, such as, there would be a Bronze players having an extremely high APM or high amount of hotkeys assigned/selected, but still stuck in the lower ranks. These players probably have the talent of understanding the game very quickly or understanding the circumstances of situations, but because they do not have the talent for playing games, are stuck in the ranks they are in. Sadly, not every person is born with greatness at every aspect of every game. It is like basketball analyzers really understand basketball and can play great ball at the high school or college level, but will never be a professional because of their body size. I believe this analogy can easily be implemented to professional e-sports gamers. However, that does not mean they do not have a place in professional e-sports. My suggestion is to create different e-sports professions: 1) E-Sports Player, 2) E-Sports Analyzer, 3) E-Sports coach and 4) E-Sports Game Tester. An E-Sports Player are students who have exceptional talent in gaming, regardless of what they play and have the ability to compete with the best in the world. E-Sports Analyzer are students who are very good at breaking down the games, the strategies and the thought process of players and analyzing them through commentator or blog posts. E-Sports coach are perfectionists who can create practice regiments and have the time to sit down to help the players become better than their current form. E-Sports Game Tester are students who understand the pros and cons of the game play, design and technical issues and can create a better game for competition. These 4 professions are just my suggestions, I am sure there are many other ways to break talent down or go into more details of the 4, but there is a lot of room to swim around and find the perfect fit for every student interested in this field.

After my analysis, I cannot help but think that the e-sports industry is vastly unexplored. Not only are there multiple genres of games with variables still no recorded, but the professions e-sports can create. This is just a stepping stone for my current state, but if new datasets for different games with even more detailed variables appear, I will definitely take the time to analyze every detail of that dataset.

If you have any questions, please e-mail me at [edwardlee0323@gmail.com](mailto:edwardlee0323@gmail.com) or find me on linkedin at [www.linkedin.com/in/leezhongpu](http://www.linkedin.com/in/leezhongpu). Thanks!