Hero Recommendation system for Dota 2

Akshay Borse
Santa Clara University
Santa Clara, CA 95050
aborse@scu.edu

Amey Baviskar
Santa Clara University
Santa Clara, CA 95050
abaviskar@scu.edu

Abstract

Recommendation system can be defined as a system which will suggest an item which has a close similarity with users which are alike. Recommendation problem can also be called as information filtering system or matrix completion problem. In this paper a recommendation engine is built around the most popular video game, Dota-2 to recommend a hero. We have also presented with insights of hero has a effect on overall win of any match. We have built upon previous work done on recommending or predicting a hero for Dota-2 and compared it with a new machine learning algorithm in search of new dimension. We have also presented exploratory data analysis on how a hero has performed in a match, feature selection and results.

1 INTRODUCTION

Video games has been in the news since Atari introduced Pong in October 1958. There has been a rapid growth in video games since the introduction of powerful GPUs and better storylines. Due to widespread internet connectivity and seamless live streaming of games players situated at any corner of the world were able to play together. DotA 2 is multiplayer online game developed by Valve. DotA was developed as an extension of World of Warcraft: The Frozen Throne game by Blizzard. Valve Corporation acquired the intellectual property rights to DotA in 2009 to develop a franchise, beginning with the stand-alone sequel, Dota 2, released in July 2013. Currently it has 10 million unique players and every year more players keep on adding. There are many tournaments held yearly when professional player compete for some prize money one of biggest tournament is DotA 2 International also known as "The International -¿ TI ". TI8 (The international 2018) was held is August 2018 and the prize pool was $25 million which is quite a significant amount considering that this just a game. [2]

DotA-2 has a only one map and the match is played between two teams total of 10 players, each team containing 5 players. Despite of being one map the complexity of the game increases due the because there are around 115 heroes and a player can choose any one hero from them to play in a match. Once a hero is chosen it cannot be used by any other player of either team. To win the match the team has to destroy a structure called the 'ancient' of the opponent team hence the name DotA(Defence of the Ancients). Every hero has its unique power, strength and weaknesses. In order to increase the strength and earn gold, a hero has to kill enemy heroes and gain experience. This gold is then used to purchase items for its hero.

The game depends on the strength of hero and also how a player can handle the hero. Assuming that all the players are of similar level, the game depends totally on hero selection. DotA-2 is in a team game, as the strengths and weaknesses of every hero has an direct effect on the team and enemy. Hero roles are divided as follows

- Position 1 : Hard Carry
- Position 2 : Solo, Ganker or Semi-Carry

- Position 3 : Offlaner or Suicide Solo
- Position 4 : Roaming Support or Jungler
- Position 5 : Hard Support or Babysitter

Despite being the role defined for every hero a professional player can choose to play a hero differently. That makes the game very complex and this is why recommending heroes make a much bigger problem than it looks. Consider in a game where the opponent is only killing or harassing a carry ignoring support this in turn giving support a chance to earn gold for himself. Now the support can carry the game or else carry for some amount of time until the main carry can get back into the game giving them a good winning chances.

1.1 Motivation

With the rapid success of Amazon, Netflix which resulted in growth in revenue and also increase in user base, we started looking for problem where it can best fit into. As video game industry is rapidly developing and Gaming now becoming a career option not just because of DotA but there are several game offering big prize money in order to attract new players. Number of players increasing and the prize money increasing per year we seeked the importance as how AI can assist a gamer to produce better results as for our case select heroes which can give a better chance of winning.

1.2 Problem Statement

We have proposed a hero recommendation system on every pick of hero based on what selections has been done by the team member which will help in improving the win probability. For example, if two out of five heroes are already chosen by a team, then the remaining three heroes are recommended on basis previous match history of user and the heroes. This recommendation will help improve the likelihood of winning. We have also presented previous studies done by [3] [5], [4], [6] and compare our results with them.

1.3 Challenges

We faced challenges in moulding the dataset which will fit into the recommendation problem. As there is no similar implementation available developing something from scratch was another challenge we faced. The data set provided on kaggle was scraped using valve api. It contains all the features related to a match. These can be used to tune the results and make the recommendation more refine.

1.4 Related Work

Many previous work are already been published with similar motive of recommending heroes, predicting who will win the match. A web application, Dotapicker [1] is developed which recommends heroes and has reported 63% accuracy of predicting the winning team. [3] proposed a recommendation system based on logistic regression having hero as feature in binary vector and k-nn to suggest distance between teams. They have reported accuracy of 68%. Another paper [6] presents a study on different approach used by different papers. [5] has focused on different features like gold per minute, experience per minute, kills per minute and various combination of these associated with heroes and used two win predictors, full post-match data and hero selection data. They have predicted accuracy of 90% using logistic regression and random forest classification.

2 DATA SET

We have used dataset provided on kaggle[6] which consist of 50,000 matches. Our data is the combination of following information.

- It consist of 50,000 matches played between 100,000 teams on an international level. The team can be same or can be different. We have 50,000 outcome of winning and losing team. The matches are complete with no interruption in between the match.

- As the role of player is also dependent so we are considering all the players on the similar skill level and keeping the game completely dependent on hero power.

- The data is stored in csv's files which we will be reading into pandas to be used with. It has data about which heroes were used to form a team, which team won, total number of death taken by a hero, gold collected, experienced earned, time taken to finish the game for every match, death, assists.

- The Dataset contains information about the matches, players, heroes, teams, teamfight, chats, hero ability. For every match, number of kill taken by each hero, when was the first blood dropped, the total number of times the hero was revived, number of assist it did while killing, etc.

3 Exploratory Analysis

3.1 Experience per minute, Gold per minute

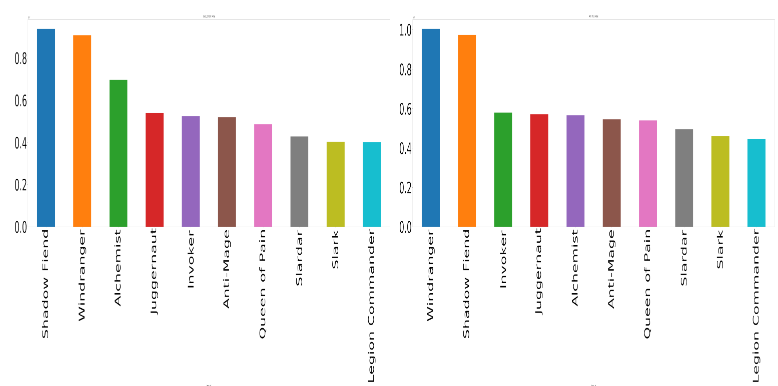
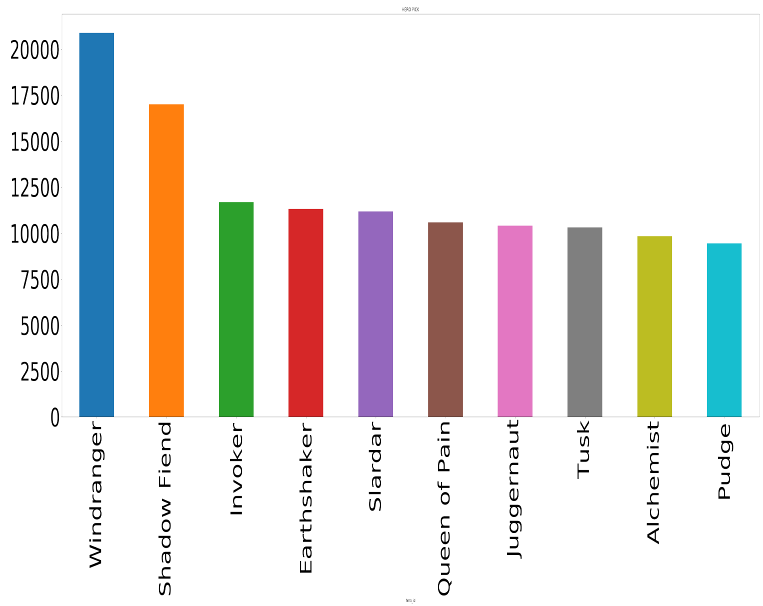


Figure 1

It can be observed from the graph that the hero who has earned more gold per minute has earned more experience, i.e. Windranger, Shadow Fiend, Invoker, Juggernaut, Alchemist thus these two features are dependant on each other.

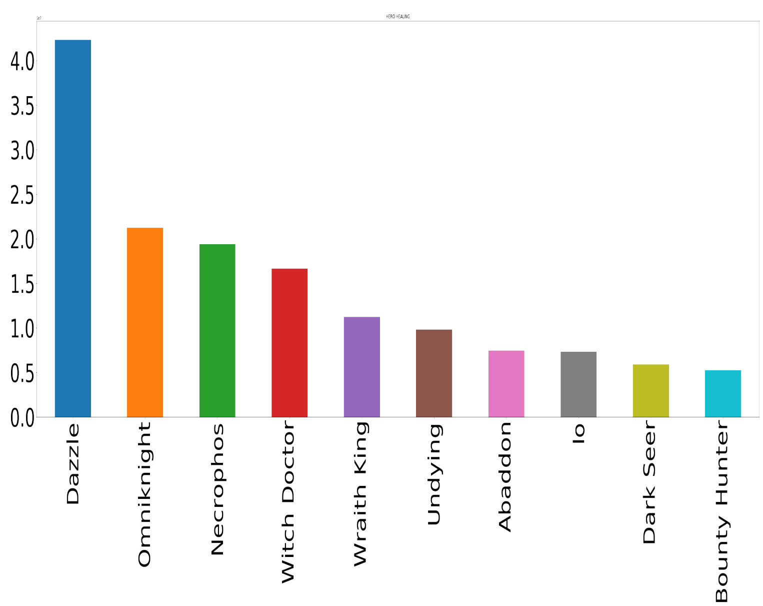
3.2 Analysis on hero picking pattern

And the hero which are mostly picked by the players are Windranger, Shadow Fiend, Invoker, Earthshaker, Slardar, Queen of Pain. Thus these heroes are mostly the choices of international players and they have win maximum matches with these.



3.3 Hero healing

It shows how a hero has helped in healing others hero. We can observed the heroes which are mentioned here are different than the heroes who gains more kills and gold. Thus it can be stated that knowing which heroes excel in which sector is very important. Dazzle, Omniknight, Necrophos, Witch Doctor, Wraith King are the top heroes who has helped healing.



4 PROPOSED SOLUTION AND CURRENT RESULTS

This section contains the solution which we are proposing compared with the previously proposed solutions. The algorithm used by us.

4.1 Feature Vector

Our algorithm works on the basis of the previous matches played and the probability of winning by each hero. There are 112 heroes in every match so our feature vector will consist of $x \epsilon \mathbb{R}^{112}$.

We have created two feature vectors. F_{win}
and F_{loss}

For F_{win}, $x \epsilon \mathbb{R}^{112}$ such that:

$$\mathrm{x}_w = \begin{cases} 0, & \text{if the hero is not chosen by a team} \\ 1, & \text{if the hero is chosen by the team} \end{cases} \tag{1}$$

For F_{loss}, $x \epsilon \mathbb{R}^{112}$ such that:

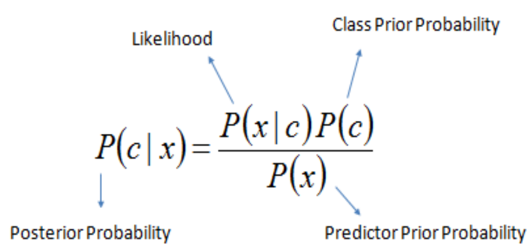
$$\mathrm{x}_l = \begin{cases} 0, & \text{if the hero is not chosen by a team} \\ 1, & \text{if the hero is chosen by the team} \end{cases} \tag{2}$$

and we will be considering previous game information as which team won on the basis of which heroes.

For the first step we are considered matches, users, heroes and there win probability. We preprocessed the data to get the required information which was located in different csv's. We further counted the wins and losses for every heroes across all the matches it has played. We then calculated win probability for each hero. We then constructed a feature matrix consist of team which has won the matches with the heroes with 1s and other 0s and the team. We trained the model by dividing the dataset into 70 and 30. Based on feature matrix thus constructed we are calculating log likelihood for each hero in a match. So when a combination of heroes is given from the testing set, it predicts rest of the heroes for the team and recommends top 10 results based on how it has performed in previous matches. Out of the humongous data we are considering the data which has teams with heroes with which they played, if they lost or won the match. So the sparse matrix thus constructed will be 50000 matches i.e. rows and 112 heroes i.e. columns. Each row will consist of five 1's means the heroes which are chosen by a team for that match and rest others will be zero. Deciding which heroes plays is an important factor.

4.2 Algorithm

The feature matrix of our model consist of the teams which have already played and hero amongst the 112 picked by team.



$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- First we calculate the prior probabilities = total teams won / (total teams won + total teams lost).
- We then calculate the likelihood of winning with particular hero = total number team won with particular hero / total number of team won.
- We didn't use evidence because it just to normalize the data our data was already normalized. Evidence = probability that a hero will be picked. Then we calculate posterior = prior * likelihood

- All the calculation were done in logarithmic base 2 as the values were very small it was easier to compare on log base.
- Considering our data set of 50000 matches the priors were same a in a match one team loses and one team wins which gives the prior probability to be 0.5
- So basically our posterior were directly dependent on the likelihoods of the heroes. Whereas the likelihood will be different for all 112 heroes.
- We then sort the posterior probability and pick the top 10 teams with highest win probability.

5 Results

We have a test data of 10K matches and we tested our recommendation system on that data. We selected each winning team and gave randomly selected two heroes out of 5 and gave it as an input the system. We ran the code to predict the remaining 3 heroes based on the the out comes of the teams for 50K matches.

As we are considering the probabilities and we have 112 heroes, around 80K teams were suggested. So we put a threshold to check for top 25 teams and verified if such team exist in testing set. Out of all the teams it predicted for 10K individual matches, we had an accuracy of 64% of predicting correct matches. Our Naive Bayes model will try to predict all the possible teams as probability is never 0. We observed that particular 5 heroes has very high probability of getting selected due to there ease of use and powers but in the real match, each team selects hero alternatively so the probability is less of acquiring a particular hero.

6 Lessons Learned

We were trying to use regular python code where as the pandas was giving faster results than the regular code. We have used only heroes as the feature for our algorithm while their attributes are also important. We learn many things from EDA like usually a team of 5 matters but there are some cases where particular combo of 2 heroes can win you the game. Having done everything by ourself we didn't realised that these algorithm can be used through default libraries.

7 Division of Work

We decide to divide each task equally so that each of us can learn and can keep a track of what is happening in the project. Data cleaning, EDA, Project report - Amey Analysis, Code, Project proposal, PPT - Akshay

8 Conclusion

The Exploratary Data Analysis helped us to understand how important is a hero in the game. Choosing the right hero at the start of the match can give an edge to the team. It is not always that a particular hero will help to change the game but the EDA showed that every hero is unique and can excel in a area. As choosing a hero has direct impact on the team, the EDA provided in this paper is helpful.

The algorithm used by us in this paper will help us to recommend the best combination of the heroes which will balance the team.

9 Github URL

`https://github.com/amsborse/Hero-Recommendation-System-for-Dota-2.git`

References

[1] Dota hero picker website. http://dotapicker.com/counterpick.

[2] Dota matches. https://www.kaggle.com/devinanzelmo/dota-2-matches.

[3] Kevin Conley and Daniel Perry. How does he saw me? a recommendation engine for picking heroes in dota 2. *Np, nd Web*, 7, 2013.

[4] Kaushik Kalyanaraman. To win or not to win? a prediction model to determine the outcome of a dota2 match. Technical report, Technical report, University of California San Diego, 2014.

[5] Nicholas Kinkade, L Jolla, and K Lim. Dota 2 win prediction. Technical report, Technical Report. tech. rep., University of California San Diego, 2015.

[6] Aleksandr Semenov, Peter Romov, Kirill Neklyudov, Daniil Yashkov, and Daniil Kireev. Applications of machine learning in dota 2: Literature review and practical knowledge sharing.”.