Data Mining in Dota 2

Samina Khan

University at Buffalo

saminakh@buffalo.edu

December 7, 2015

Overview

Introduction

Dota 2 is a multiplayer online game, where matches are played between two teams of five players controlling their own hero chosen at the start of the match. Each hero has a unique set of skills, starting attributes and attribute gain rates, resulting in different strengths and weaknesses for a hero. As a result, match outcomes are not only dependent on the mechanical abilities of a player, but on team composition and hero synergy as well.

Introduction

Problem

By looking at the team composition of the winning teams in matches, we can identify pairs of heroes that synergize well together. We can also identify "counterpicks," where the choosing of a hero significantly reduces the viability of a hero on the opposite team.

Combos	**Counterpicks**
Bane + Mirana	Dazzle + Legion Commander
Tiny + Wisp	Doom + Storm Spirit

Table: Hero pick and counterpick examples

Introduction

Impact

With this project, we are aiming to identify successful combo picks as well as popular counter picks in the current Dota 2 meta. This could allow players to

- learn of popular combos and counter picks in higher skill brackets
- learn how to anticipate and play around them

Formulation

Task

Teams in Dota 2 are composed of five distinct heroes each, selected from a pool of over 100 heroes. Identifying pairs of heroes often picked or counterpicked together is similar to finding items found together frequently in a transaction. Thus, the apriori algorithm was expected to suit the task.

- Input: list of matches, with each match indicating the heroes selected
- Output: list of heroes picked together above the required threshold

Datasets

Valve Web API

Dota 2 public matches are available in .json format, which can be obtained by registering an API key with Valve. The .json returned includes

- The match ID
- A .json array of the player accounts and their selected hero
- Lobby type (ie 1v1, Ranked, etc.)

Queries

There are also options that can be included in the query to limit the results, such as

- skill= (0 is any skill bracket, 0 is normal, 1 is high, 2 is very high)
- start_at_match_id= (allows you to start the list list of matches at a particular one)

Datasets

Collecting Data

I chose to use Java to periodically query the Valve servers and save the relevant matches to a text file. There were multiple things that had to be taken into consideration when deciding what matches to use:

- Only using matches from high and very high skill brackets - Dota 2 has a bot problem with farming items in lower brackets. By only looking at matches in high and very high brackets, we avoid including bot matches in the data
- Only using ranked matches - this filters out 1v1 matches and custom matches and maps
- Discarding matches with any player having a hero ID of 0 - a 0 indicates a player that abandoned before choosing a hero

Algorithm

Apriori Algorithm

I decided to use the apriori algorithm to analyze the data, as selecting heroes and finding rules to represent heroes commonly selected together is very similar to the process of items selected together in a transaction.

- Each hero represented an 'item' out of a pool of 112 heroes
- Each team was treated as a single transaction in the case of determining combos, while the enemy and ally team were considered a single transaction for counterpicks
- For combos, this was about 6000 high to very high skill bracket matches, with 3000 for counterpicks
- The numbers returned are the hero id (an int from 1-112) which can be looked up using Valve's web api

Experiments

Combos

Running the apriori algorithm on the combos dataset yielded the results below. The picks were dominated by Shadow Fiend, Juggernaut, Pudge, Invoker, Windranger and Emberspirit. In addition, Anti-mage, Earthshaker, Bounty Hunter and Tusk were very popular.

1	7	8	8	8	11	11	11	11	11	14	14	21	21
74	11	11	21	74	14	21	62	100	106	21	74	74	106

Table: Counterpick L2

Experiments

Counterpicks

Running the apriori algorithm on the counterpicks dataset yielded the results below; as with the combo picks, the counterpick heroes were primarily Shadow Fiend, Juggernaut, Pudge, Windranger, QoP, Invoker, and Ember Spirit.

11	11	11	11	11	11	14	21	74	74
8	14	21	39	74	106	21	11	11	14

Table: Counterpick L2

Experiments

Results

Counter to my expectations, it appears that in the high and very high skill brackets for ranked matchmaking, players do not attempt to combo with their allies or counterpick their enemies. Rather, they pick heroes that are popular in the current meta and rarely seem to deviate from those standard picks.

Experiments

Potential Explanations

These results were intially surprising, but are certainly a reasonable outcome given the selected data.

- Perhaps non-ranked matches would have resulted in a larger variety of picks, where there is less at stake (and thus more experimentation in strategy and combos)
- A lower skill bracket, assuming bot matches were detected and removed from the sample data, may also have a larger variety of picks where "sticking to the meta" isn't as common
- Maybe balancing in the game has gotten to the point where cheesy combos or explicit counterpicks just aren't as successful as they used to be

Challenges

Valve Web API

- Some queries did not work correctly (such as match mode, which would have allowed me to use non-ranked while still avoiding 1v1s)
- Json data returned was not complete, and to get additional match information would require more queries
- Only returns the most recent 25 matches, have to use workaround to get later matches (and it dosen't work very well)

Personal challenges

I am not proficient with MySQL or any other database software (other than Realm, which is Android-only). Having database knowledge for saving match history would have helped immensely, but because I did not, I had to save the matches to a text file and load a hashset of (int, int[]) pairs of the match ID and hero IDs each time I opened the program to avoid adding duplicate matches.

Links

- Github for code: `https://github.com/saminakh/CSE469`
- Dota 2 dev forums for web api: `http://dev.dota2.com/showthread.php?t=58317`
- Github for interesting machine learning project and paper on a Dota 2 counterpicking tool: `https://github.com/kevincon`