Role Identification in Dota 2 for More Effective Analysis of Player Performance

Introduction

Esports games generate a vast treasure trove of data. This data is of value both to the players competing for multi-million-dollar prize money, and to the broadcasters who want to satisfy the hundreds of millions of viewers who, like fans of traditional sports, have an immense appetite for accurate gameplay statistics and analysis.

One of the most popular esports formats is the Multiplayer Online Battle Arena (MOBA), in which two teams of players face off in an attempt to destroy the opponents base. Each player controls a particular hero, and carries out different roles or functions to help achieve victory for their team.

To understand how well a particular player has performed, it is necessary to understand the role of their hero within the team. Unfortunately, in esports, roles on a team are less easy to identify than they are in physical team sports; even experts can sometimes disagree about the specific role that a hero is playing.

To complicate matters further, the role assigned to a player's hero can switch between games, which means that the same metrics are not always appropriate for evaluating their performance. In a traditional team sport such as football, it would be the equivalent of using goals scored for the team as a measure for judging a player picked to play as striker one week and then goalkeeper the next.

The net result is that esports analytics systems have historically struggled to identify the role being played by a hero in a particular game. This has made it difficult to identify the correct metrics by which to judge their performance.

In 2018 and 2019, we undertook a study that sought to solve this problem by making innovative use of cluster analysis to identify a player's role correctly and build an appropriately labelled dataset to allow for accurate performance analysis.

Concentrating on Dota 2, one of the most played esports titles in the world, the study has laid the foundations for more effective analysis of player performance in a range of MOBA titles as well as other esports.

The here and now: Existing role identification

Existing methods of role identification in MOBAs rely on a combination of two approaches. One is to label the data based on the opinion of esports experts, which introduces the possibility of bias and involves a highly laborious process of manual data annotation. It is also dependent on preconceived, but untested ideas about how player roles should be defined.

The other is to identify a role by making use of seemingly objective metrics. The problem here is that many of the metrics used are too heavily skewed by the overall result of the game. In almost any kind of team game, it is possible to perform excellently but still fail, perhaps because your opponent performed better, or your teammates let you down, or you had too few opportunities to make a difference in your particular role. Relying on metrics that depend too heavily on the outcome of the game can therefore lead not only to the inaccuracies of measuring a player's performance but even to the mis-identification of their role.

Our study was designed to address the flaws in both these approaches by building towards a new methodology of role identification, one capable of delivering a labelled dataset without relying on misleading metrics or subjective opinion.

The future: Improved datasets for an enhanced understanding of esports at all levels

Our starting point was to define a set of features for identification that are agnostic to player/team performance and of the game's outcome.

We identified three features that are indicative of a player's role in Dota 2: Map Movement (where a player spends their time during the first 10 minutes of the game); Resource Priority (how much of the team's resources are given to a player); and Ability Prioritisation (the order in which a player chooses to build up the abilities of their hero). We then used three distinct types of clustering algorithms to analyse these groups of features, and compared the results to determine which one yielded the best results for the feature.

Our next step was to take the chosen cluster for each group and combine them using ensemble clustering. This gave us a precise clustering across all features, showing how each of the three groups correlate with each other. From here, we created role classifications that are defined by the data rather than by human preconceptions.

Finally, we used these data-defined classifications to create a large automated role-labelled dataset. This dataset can be used to identify the specific features or flaws of each hero character in the numerous roles that they could perform.

Our chosen dataset, retrieved from OpenDota, was made up of the last 5,000 professional and semi-professional Dota 2 games played prior to 5 December 2018. More specifically games played at Dota 2 major and minor tournaments.

As a result we were able to devise a novel method of role detection that is unbiased by performance, and demonstrate its application by creating a large, automated and accurately labelled dataset.

Summary and next steps

We believe that our work represents a significant improvement to the accuracy, objectivity and automation of dataset creation. This in turn will increase the accuracy of the tools currently used to analyse such datasets, such as the prediction model tools that are applied to data collected from heroes within a game.

Our method of role identification could also be applied to the existing historical data for each hero, thereby improving our understanding of how the different heroes tend to perform in particular roles.

This approach is not limited to Dota 2 and could be readily applied to various team-based esports. Most obviously for similar MOBAs such as League of Legends, which make use of the same three features although with different names, and for team-based first-person shooter games such as Counter Strike Global Offensive where the concept of roles is equally important.

It would also be possible, when adapting our methodology for a specific MOBA or other form of esports, to add further features to the analytics framework, such as 'Runes' for League of Legends, or 'Loadout' for Counter Strike.

Although our approach has focused primarily on role identification in professional and semi-professional gameplay, it could readily be applied to lower-skilled data. It is likely that less-skilled players would not stick to their assigned roles as closely, which will result in a broader, less defined set of clusters. However, determining the difference between 'amateur' and 'professional' clusters could have valuable applications in data-driven coaching, by helping players to learn the 'proper' roles more effectively.

There are multiple other use cases for this work: professional post-game review; casual player analysis; and improvement to the accuracy of existing analytical tools. The research could also help to expand the accessibility of esports by enabling broadcasters to present more insightful live game analysis.

Link to academic paper: https://ojs.aaai.org/index.php/AIIDE/article/view/5235/5091