

Diseño de un modelo predictivo para equipos profesionales en Dota 2

Anteproyecto

Línea de investigación del grupo FICB-PG: Línea de Investigación en educación y tecnología

Rene Felipe Cardozo 0321040262

Asesor: Javier Niño

Diciembre 2018

Resumen

La investigación de en esta tesis consiste en explorar los datos de las partidas del juego Dota 2, donde se analizarán las partidas singulares de los jugadores pertenecientes a un equipo profesional y en las partidas realizadas en ligas mayores y premier. Por un lado, analizaremos las partidas singulares de un jugador profesional, entrenando algunos algoritmos, para obtener un modelo predictivo. Por otro lado, se implementará un árbol de decisión que será a su vez alimentado por el anterior modelo aplicando este árbol en el campo de las ligas mayores y premier, para predecir un equipo ganador.

Palabras clave: Machine learning, modelo predictivo, minería de datos, aprendizaje supervisado.

Abstract

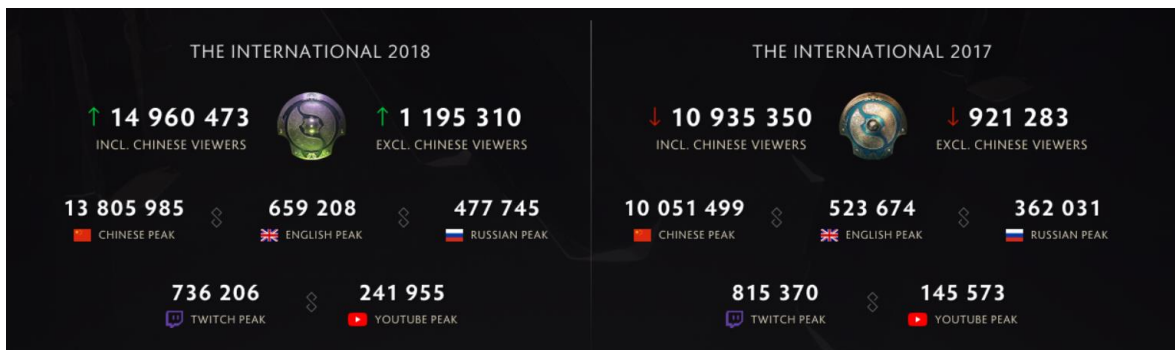
The research that will be carried out in this thesis will consist in exploring the data of the games of the game Dota 2, where the unique games of the players belonging to a professional league will be found, and in the games made in the major and premier leagues. On the one hand, we will analyze the unique games of a professional player, training a series of supervised learning algorithms and obtain a predictive model. On the other hand, a decision tree will be implemented that will be fed by the previous model, applying this tree in the field of the major and premier leagues, to predict a winning team.

Key words: Machine learning, predictive model, data mining, supervised learning.

INTRODUCCIÓN

Según Wagner la definición de sport es “un área de actividades deportivas en la que las personas desarrollan y entrenan habilidades mentales o físicas en el uso de las tecnologías de la información y la comunicación” [1].

Abro esta introducción con la definición dada por Wagner sobre esports, aunque muchas veces considerados como un entretenimiento. Existen varias estadísticas que demuestran que los esports ya han alcanzado grandes audiencias, como es el caso de League of legends con 14.7 millones de espectadores recurrentes, teniendo en cuenta que el juego de 7 de la NBA en el 2016 fue de 44.5 millones de espectadores [2]. Una cifra considerable teniendo en cuenta que la de league of legends o su juego rival Dota 2 sigue creciendo.



[3]

¿Qué es Dota 2?

Dota 2 es un juego gratuito **Multiplayer Online battle arena (MOBA)** desarrollado por la corporación Valve. Consta de dos equipos llamados Dire y Radiant, cada equipo se conforma por 5 héroes diferentes, gana el equipo que destruya el edificio “ancient” del otro equipo. No existe un límite de tiempo para lograr destruir el edificio objetivo “ancient” y su mapa al igual que muchos otros MOBA como *League of Legends* o *Heroes of the Storm*, se basa en tres líneas protegidas con torres y para destruir el último edificio es necesario primero destruir estas torres en las diferentes líneas, existen cerca de 215 héroes elegibles, cada uno con diferentes poderes y habilidades.



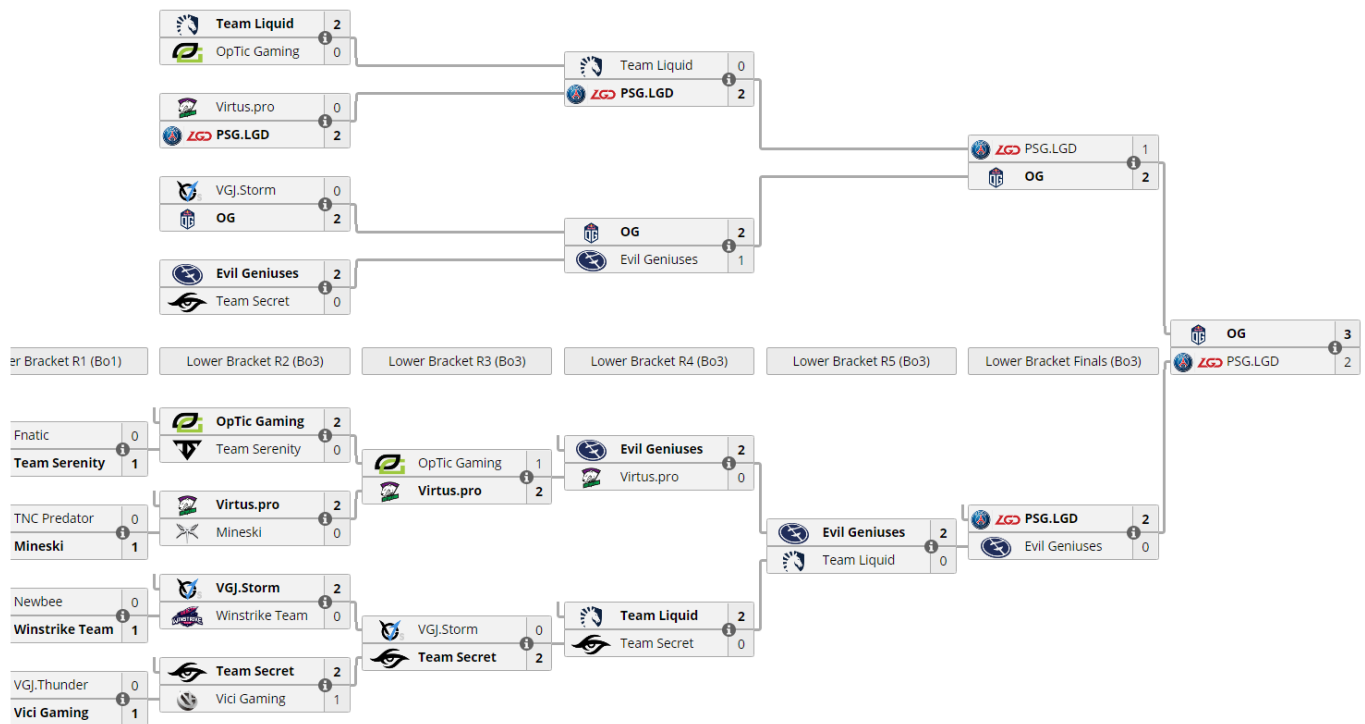
[4]

Equipos profesionales

En Dota 2 existen ligas de diferentes tipos, existen las más grandes que son la premier y las ligas mayores, en las cuales los equipos élite compiten para llevarse un gran premio, su liga de mayor representación es “The International” donde el premio a obtener como primer lugar es de 11'190.158 USD [5]. Existen otras ligas donde fácilmente el primer lugar puede ganar 350.000 USD [6].

En la actualidad los equipos profesionales compiten en una liga mayor llamada “The International” que es llevada a cabo cada año desde el 2012. Esta competencia tiene un modo para los usuarios normales llamado el compendio donde aquellos usuarios que compran un pase de batalla en la temporada de la liga tienen derecho a votar por su equipo favorito. Esta característica tiene la posibilidad de completar en un árbol de posibles equipos ganadores hasta predecir qué equipo puede ser el ganador de esta liga [7].

El juego en este punto es muy versátil, cada liga tiene unos brackets muy particulares, es decir, las fases son muy parecidas a un mundial de Fútbol, tiene clasificados los mejores equipos, y su apertura inicia con una nueva combinación de equipos para cada nueva versión de la liga.



[8]

Se realizan diferentes tipos de ligas en Dota 2: ligas premier, ligas mayores y ligas menores y algunas líneas amateurs que son online. Al año, se juegan alrededor de 10 ligas premier [9], 17 ligas mayores [10], cada una de estas partidas se puede asistir presencialmente o ser espectador por algún canal de streaming como twitch.tv para ver cada partida. En un día, esta transmisión puede llegar a tener un pico de 14 millones de espectadores simultáneos observando este tipo de eventos [11].

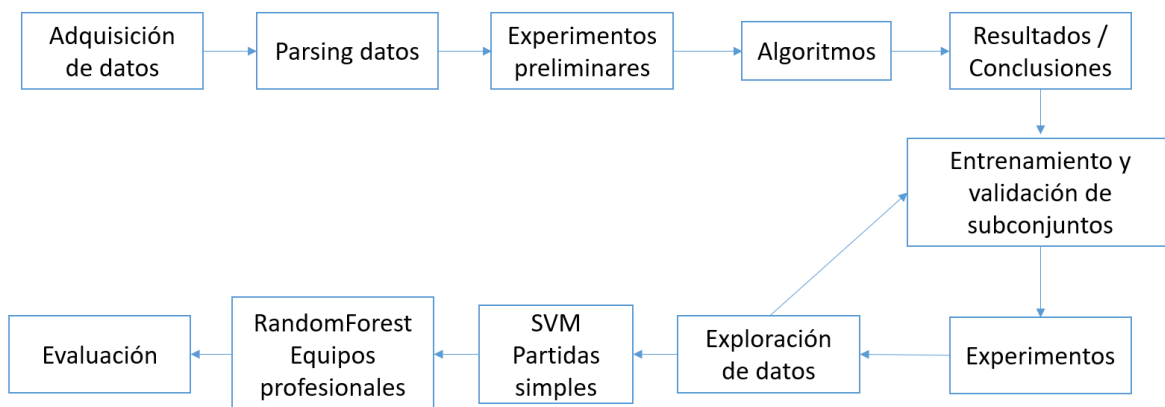
Machine Learning

Machine learning es un amplio campo de las ciencias de la computación que gobiernan sistemas que pueden aprender de un conjunto de datos en orden de mejorar el rendimiento de una tarea específica [12]. En este campo se encuentra una subsección de problemas llamado aprendizaje supervisado, donde la entrada es usada para crear un modelo el cual produce una salida deseada [12]. Si vemos los métodos de aprendizaje como una caja negra, decimos que la entrada son un conjunto de instancias donde cada instancia tiene un vector de atributos, uno de esos siendo la salida deseada.

Los modelos son contruidos para emparejar las relaciones desde los atributos a la etiqueta a predecir. Si este atributo es un valor nominal, es llamada la clase y el modelo un clasificador, allí es donde hay bastante variedad de algoritmos de aprendizaje debido a los diferentes modelos de operación, para construir modelos que funcionan internamente diferente, es decir cada uno tiene su propia implementación si lo vemos desde el punto de vista estratégico.

METODOLOGÍA

La metodología que ya se encuentra en progreso en este proyecto de investigación se presenta a continuación, seguida de su significado en cada etapa.



Adquisición de datos: En esta etapa se presenta el API y se construye un software capaz de obtener los datos por cada jugador profesional que activamente pertenezca a un equipo de una liga profesional. Este proceso se encuentra en progreso, en la actualidad para la construcción del microservicio se eligió Spring Boot por su facilidad y adaptabilidad a llamados REST, almacenando todos estos llamados en una Base de datos mongoDB.

Parsing data: Esta etapa se trata de garantizar la homogenización de datos encontrados a lo largo de los documentos json, dependiendo de lo que requiera los diferentes tipos de algoritmos que se pretendan implementar.

Experimentos preliminares: En esta etapa se generará un primer conjunto de entrenamiento obtenido del historial de partidas jugadas de los jugadores profesionales seleccionados.

Algoritmos: En esta etapa, a través del conjunto de entrenamiento se harán pruebas de algoritmos de aprendizaje supervisado, donde obtendremos una perspicacia inicial sobre la exactitud, sensibilidad y especificidad.

Resultados-Conclusiones: Aquí nos centraremos en analizar y concluir las etapas primarias de la metodología, que mejoras se pueden hacer en los datos capturados, que algoritmos y métodos podríamos potencialmente elegir.

Entrenamiento y validación de subconjuntos: Esta etapa particularmente se hará validación de un conjunto de datos más grande, teniendo en cuenta lo aprendido en las etapas anteriores y aplicar los modelos a los algoritmos elegidos.

Experimentos: Esta etapa será fundamental para aplicar los modelos y conjuntos de entrenamiento y de prueba elegidos en etapas previas.

Exploración de datos: Aquí se presenta las últimas etapas de prueba y error, dando la posibilidad de ser necesario retroceder un par de etapas para garantizar que tanto, el conjunto de datos, como las características y los atributos son los ideales para el modelo predictivo.

SVM para partidas simples: En el conjunto de prueba se aplicará el modelo predictivo para el historial de partidas singulares de los jugadores profesionales.

Random Forest para equipos profesionales: Esta etapa, es importante porque se alimenta consecuentemente de la etapa previa, para poder construir el árbol de entrenamiento y poder aplicar el conjunto de prueba.

Evaluación: Finalmente después obtener los datos y la información interpretada, podremos evaluar todos estos factores y poder llevar todo esto al conocimiento y generar un análisis basado en la experiencia de este ciclo metodológico.

Objetivos

Los objetivos son la pieza clave, en esta sección presento el objetivo general y los objetivos específicos.

Objetivo General

Diseñar un modelo predictivo usando máquinas de aprendizaje supervisado que prediga el equipo ganador en un torneo o liga de Dota 2 dado los datos parciales recopilados de ligas anteriores y partidas comunes de cada jugador profesional que participe en estas ligas.

Objetivos Específicos

- Determinar cuál es la relación entre las partidas que juega un jugador profesional fuera de la liga y el rendimiento de ese jugador cuando participa en una liga profesional.

- Proponer un modelo predictivo de que un equipo pueda o no ganar una partida en una liga teniendo en cuenta el historial de ligas.
- Elaborar un modelo en el cual se analicen los resultados singulares de los jugadores profesionales en sus partidas fuera de campeonato.

Preguntas de investigación

P1: ¿Cuál es la precisión de nuestro modelo predictivo para determinar si un equipo puede o no ganar una partida en una liga?

P2: ¿Cuál es la relación entre las partidas singulares y el rendimiento de un jugador en medio de una liga?

P3: ¿Cuál es la precisión más alta alcanzable usando diferentes atributos en la ejecución de Random Forest?

TRABAJO REALIZADO – ESTADO DEL ARTE

On using Artificial Neural Network models to predict game outcomes in Dota 2

Esta tesis se implementa redes neuronales y exploran con este método la selección de personajes o héroes del juego, este modelo considera únicamente dicha selección y hace una variación de neuronas para analizar su comportamiento a medida que aumentan las capas, dando a entender que es proporcional al número de predicción llegando a un 59% [13].

Result Prediction by Mining Replays in Dota 2

Esta tesis, idea un modelo de aprendizaje basado en el estado en diferentes tiempos durante un juego, predice un equipo ganador dado estos diferentes estados, implementando Random Forest después de hacer pruebas con otros algoritmos concluyendo porqué RF fue el mejor candidato para predecir la victoria sobre una partida [14].

Outcome Prediction of DOTA2 Based on Naïve Bayes Classifier

Aunque los supuestos independientes son a menudo inexactos, de hecho, algunas de las propiedades del clasificador Naive Bayes lo hacen sorprendentemente efectivo en la práctica. El autor proporciona una forma de analizar las alineaciones y la probabilidad de ganar en el Dota2 con el clasificador Naive Bayes, presenta la

idea básica de cómo analizar el juego del clasificador Naive Bayes y verifica la posibilidad de analizar el juego con datos cuantitativos en el modelo de clasificador Naive Bayes [15].

Predicting Future States in DotA 2 using Value-split Models of Time Series Attribute Data

En este trabajo, se introduce un enfoque para pronosticar cambios en la salud del héroe en DotA 2 al dividir los datos en cambios grandes y pequeños, usando este enfoque de división de valores, se predice ambos tipos de cambios por separado utilizando diferentes modelos estadísticos. Para cambios pequeños, usan un modelo de media móvil autorregresiva (ARMA) y para cambios grandes usan una combinación de métodos. Los cambios grandes ("puntos de salto") se predijeron utilizando una estimación no homogénea del proceso de puntos de Poisson, mientras que la regresión logística y la regresión lineal se usaron para predecir el signo y la magnitud de estos puntos, respectivamente [16].

Prediction of Dota 2 Game Result

La parte teórica de esta tesis se centra en aclarar brevemente el árbol de decisiones y la teoría de redes neuronales artificiales, explica los factores básicos que tienen un impacto significativo en el resultado del juego. En la parte práctica, el enfoque se centra en experimentar con los parámetros de la técnica de aprendizaje automático, extender los datos de entrada con información sobre las composiciones de héroes, comparar y evaluar el rendimiento de estas extensiones. Todo esto da como resultado la implementación de un programa experimental que producirá un modelo ANN predictivo. Este modelo se puede usar más adelante para predecir el resultado del juego según el conocimiento de las composiciones del héroe del equipo inicial [17].

Bibliography

- [1] M. G. Wagner, "On the Scientific Relevance of eSports," in *Proceedings of the 2006 International Conference on Internet Computing & Conference on Computer Games Development*, Las Vegas, Nevada, USA, 2006.
- [2] K. Noonan, "The Motley Fool: Stock Investing Advice | Stock Research," The Motley Fool, 26 11 2017. [Online]. Available: <https://www.fool.com/investing/2017/10/25/7-gaming-stats-that-prove-esports-is-the-next-big.aspx>. [Accessed 30 11 2018].
- [3] esc.watch, "Researching esports and streaming trends," 27 08 2018. [Online]. Available: https://esc.watch/storage/app/media/uploaded-files/TI_All_Stages.png. [Accessed 04 12 2018].
- [4] Research Gate, "Discover scientific knowledge, and make your research visible.," researchgate.net, 2018. [Online]. Available: https://www.researchgate.net/figure/Map-of-Dota-2-from-Dota-2-wiki-7_fig1_262207918. [Accessed 21 11].
- [5] V. Corp, "Dota 2 - The International," Valve, 20 08 2018. [Online]. Available: <http://www.dota2.com/international/overview/>. [Accessed 06 11 2018].
- [6] Liquid, "Liquipedia Kuala Lumpur Major," Team Liquid, 2018. [Online]. Available: https://liquipedia.net/dota2/PGL/Kuala_Lumpur_Major. [Accessed 06 11 2018].
- [7] Dota2, "Battlepass TI," Dota 2, 2018. [Online]. Available: <https://www.dota2.com/international/battlepass/>. [Accessed 08 11 2018].
- [8] Liquipedia, "Liquipedia," Liquipedia, 2018. [Online]. Available: https://liquipedia.net/dota2/The_International/2018. [Accessed 20 11 2018].
- [9] T. Liquid, "Premier tournaments," Liquid, 2018. [Online]. Available: https://liquipedia.net/dota2/Premier_Tournaments. [Accessed 05 11 2018].
- [10] t. Liquid, "Major tournaments," Team Liquid, 2018. [Online]. Available: https://liquipedia.net/dota2/Major_Tournaments. [Accessed 05 11 2018].
- [11] "Researching esports and streaming trends," ESM.one, 2018. [Online]. Available: <https://esc.watch/blog/post/stats-international-2018>. [Accessed 05 11 2018].
- [12] M. M, R. A and T. and A, *Foundations of Machine learning*, London, England: The MIT Press, 2012.
- [13] W. VIKTOR and A. JULIEN, *On using Artificial Neural Network models to predict game outcomes in Dota 2*, Stockholm, Sweden, 2017.

- [14] J. W. Filip Johansson, Result Prediction by Mining Replays in Dota 2, Karlskrona, 2015.
- [15] K. Wang and W. Shang, Outcome Prediction of DOTA2 Based on Naive Bayes Classifier, Beijing, 2017.
- [16] Z. Cleghern, O. Ozaltın, S. Lahiri and D. Roberts, Predicting Future States in DotA 2 using Value-split Models of Time Series Attribute Data, North Carolina, 2017.
- [17] F. Beskyd, Result, Prediction of Dota 2 Game, Prague, 2018.