

# Uma métrica interpretável para jogadores de DOTA2

Gabriel Vita Silva Franco

Departamento de Informática

Universidade Federal de Viçosa

Viçosa, Brazil

[gabrielvsfranco@outlook.com](mailto:gabrielvsfranco@outlook.com)

Giovanni Ventorim Comarela

Departamento de Informática

Universidade Federal de Viçosa

Viçosa, Brazil

[gcom@ufv.br](mailto:gcom@ufv.br)

Marcos Henrique Fonseca Ribeiro

Departamento de Informática

Universidade Federal de Viçosa

Viçosa, Brazil

[mhribeiro@gmail.com](mailto:mhribeiro@gmail.com)

**Resumo**—DOTA2 é um dos maiores *e-sports* do mundo, mas que não tem uma métrica embasada para avaliar jogadores. Com isso, o objetivo desse estudo é propor uma nova forma de avaliar jogadores de DOTA2, os quais possuem diferentes perfis. Para isso, é necessário criar uma base de dados de partidas de DOTA2, descobrir perfis de jogadores diferentes nesses dados e escolher os atributos que melhor avaliam esses diferentes perfis. Segundo essa metodologia, foi proposta uma nova métrica, chamada ADG, que consegue separar melhor os grupos distintos encontrados nos dados.

**Keywords**-mineração de dados; DOTA2; seleção de atributos;

## I. INTRODUÇÃO

Os *e-sports*, como são chamados os jogos eletrônicos que possuem campeonatos, estão crescendo de uma forma absurda nos últimos anos. Os *e-sports* movimentaram 1,5 bilhão de dólares em 2017 [1]. Os jogos do gênero MOBA (*Multiplayer Online Battle Arena*) são um dos maiores destaques nesse grande crescimento dos *e-sports*. Em jogos de MOBA, um jogador controla um único personagem em uma equipe, e estes competem contra outro time de jogadores. O objetivo principal do jogo é destruir a estrutura principal do time adversário, com a ajuda de unidades controladas por computador, que são geradas periodicamente e que se movem ao longo de trilhas definidas.

DOTA2 é um dos jogos de MOBA mais famosos da atualidade. Os campeonatos de DOTA2 distribuíram 138 milhões de dólares apenas em premiações desde 2012 [2]. Na última edição do maior campeonato anual do jogo, chamado *The International*, que foi realizado em 2017, foram distribuídos 24,7 milhões de dólares em premiação, segundo o site DOTA2 Prize Pool Track [3].

No DOTA2, cada jogador desempenha uma função diferente no time. Cada uma dessas funções acaba priorizando atributos diferentes, e consequentemente, jogadores com diferentes funções possuem bons atributos distintos. Vários desses atributos podem ser utilizados para caracterizar a função que o jogador desempenha e o quanto bom ele é. Os especialistas do jogo sabem fazer várias análises utilizando esses atributos, mas não existe uma métrica embasada para avaliar um jogador, levando em consideração as diferentes funções existentes.

O objetivo desse trabalho é criar uma métrica que consiga avaliar jogadores com diferentes perfis da forma mais justa possível. Para isso, foi utilizado aprendizado não supervisionado para encontrar os diferentes grupos de jogadores existentes. Após encontrar esses grupos, foi feita uma seleção de atributos, afim de descobrir quais os melhores atributos para compor a métrica. Foram propostas três métricas com os melhores subconjuntos de atributos, e dessas três, foi escolhida a que conseguiu representar melhor as diferenças entre os grupos.

## II. MOTIVAÇÃO

DOTA2 é um jogo de alcance mundial, com milhares de jogadores, mas que não tem uma forma bem embasada de avaliá-los de acordo com os diferentes perfis existentes. Grandes sites de estatística do jogo, como o DOTABUFF<sup>1</sup>, utilizam uma métrica chamada KDA, que é dada por:

$$KDA = \frac{kills + assists}{deaths}$$

A hipótese do trabalho é que essa métrica não é suficiente para representar todos os tipos de jogadores que existem no DOTA2. Existem outros atributos que podem trazer uma riqueza maior de informações, e consequentemente, gerar uma métrica melhor. Um exemplo disso está mostrado na Figura 1. Os Jogadores 1 e 2 são da base de dados montada para esse estudo. Essa base de dados será detalhada na Seção III. O Jogador 1 possui o KDA normalizado de 1,548357, e o Jogador 2 de 0,954751. O Jogador 2 têm alguns atributos bem superiores ao Jogador 1, e provavelmente é um bom jogador em uma função diferente do Jogador 1. Como o KDA não leva em consideração esses outros atributos, o Jogador 2 é bem inferior ao Jogador 1 em função do KDA.

O que este trabalho busca é uma métrica que consiga capturar essas diferenças de funções, e avaliar jogadores de uma forma melhor e mais justa. Uma boa métrica precisa ser interpretável, simples e representativa. Fazer uma métrica que satisfaça esses requisitos é um grande desafio, e é isso que este trabalho visa alcançar.

<sup>1</sup><https://www.dotabuff.com/>

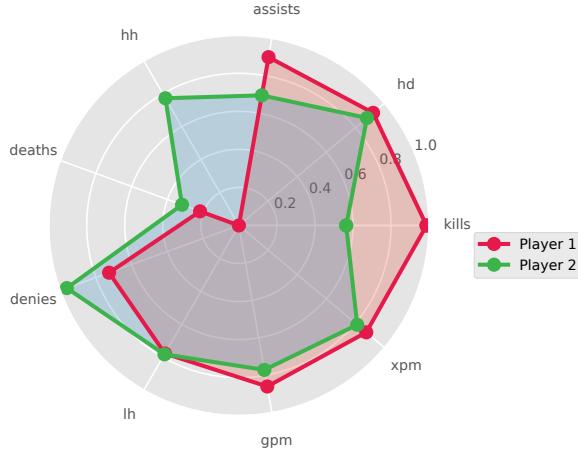


Figura 1. Comparação entre dois jogadores da base de dados

### III. COLETA E PRÉ-PROCESSAMENTO

Para a base de dados ser criada, foi utilizada a *API* da Valve<sup>2</sup>. Foram coletadas partidas de ligas suportadas pelo DotaTV, realizadas no período de 2012 a 2016, e que possuíam pelo menos vinte partidas jogadas por dez jogadores humanos. As partidas mais recentes fornecem um número maior de atributos comparado as mais antigas. Por esse motivo, foi escolhido um subconjunto que é comum em todas as partidas e que representa dados do jogador, já que este é o objeto de análise. Os dados possuem, no total, 105.662 partidas de 1641 ligas.

Tabela I  
ATRIBUTOS DA BASE DE DADOS

Atributos	Significado
Kills	Abates de jogadores inimigos
Deaths	Número de mortes
Assists	Assistências em abates de jogadores inimigos
Denies	Unidades inimigas negadas
Gold per minute (gpm)	Taxa de ouro obtida pelo número de minutos jogados
Hero Damage (hd)	Quantidade de dano causado em jogadores inimigos
Hero Healing (hh)	Quantidade de cura em jogadores aliados
Last Hits (lh)	Unidades inimigas abatidas
XP per minute (xpm)	Taxa de experiência obtida pelo número de minutos jogados

Após a criação dos dados, foi aplicado um filtro para retirar jogadores com menos que 5 partidas. Também foram feitas duas normalizações nos dados: pelo número de partidas, que consiste em dividir todos os atributos de cada jogador pelo número de partidas do mesmo, e a min-max, deixando todos os atributos no intervalo [0.0, 1.0]. Os dados

estão sendo representados em uma matriz  $D$  onde as colunas são os atributos e cada linha é um jogador. Essa base não possui itens faltantes, ou seja, todos os jogadores têm valores para todos os atributos.

Foi feita a remoção de jogadores discrepantes da base de dados utilizando o diagrama de caixas. Um jogador é considerado discrepante se o valor de pelo menos um dos seus atributo for menor que  $Q1 - 1.5IIQ$  ou maior que  $Q3 + 1.5IIQ$ , sendo IIQ o intervalo interquartil e Q1 e Q3 o primeiro e terceiro quartis, respectivamente. Esse procedimento retirou 2862 jogadores dos dados, deixando 17682 no total.

É necessário descobrir se existem grupos de jogadores bem definidos, e quantos estes são, já que os dados não têm classe. Isso sugere a aplicação de técnicas de agrupamento, e seu uso e resultados serão descritos melhor na próxima seção.

### IV. AGRUPAMENTOS

No DOTA2 existem jogadores com funções e estilo de jogo bem diferentes. Isso deixa claro que existem grupos de jogadores, mas não mostra quantos e o quão eles são distintos. Para descobri-los, será utilizada uma técnica de agrupamento, mais especificamente o *k-means* [4], como mostrado no Algoritmo 1 [5]. As decisões tomadas em relação ao número de grupos utilizado no algoritmo, a análise da qualidade do agrupamento e a comparação dos grupos com a métrica utilizada no jogo serão mostradas nas subseções a seguir.

---

#### Algoritmo 1: K-means

---

**K-means ( $D, k, \epsilon$ ):**

$t = 0$

Inicialize aleatoriamente  $k$  centróides:

$$\mu_1^t, \mu_2^t, \dots, \mu_k^t \in \mathbb{R}^d$$

**repete**

$$t \leftarrow t - 1$$

$$C_j \leftarrow \emptyset \text{ for all } j = 1, \dots, k$$

**para cada**  $x_j \in D$  **faça**

$$j^* \leftarrow \operatorname{argmin}_i \{\|x_j - \mu_i^{t-1}\|^2\}$$

$$C_{j^*} \leftarrow C_{j^*} \cup \{x_j\}$$

**fim**

**para cada**  $i = 1$  **até**  $k$  **faça**

$$u_i^t \leftarrow \frac{1}{\|C_i\|} \sum_{x_j \in C_i} x_j$$

**fim**

$$\text{até } \sum_{i=1}^k \|\mu_i^t - \mu_i^{t-1}\|^2 \leq \epsilon;$$


---

#### A. Escolha do número de grupos

Para escolher o número de grupos  $k$  do *k-means*, foi utilizada a seguinte heurística: dada a função da inércia pelo número de grupos, o  $k$  será o ponto onde essa função muda

<sup>2</sup><https://wiki.teamfortress.com/wiki/WebAPI>

de comportamento. Com base nesta análise, foi escolhido o valor de 10 para o número de grupos. O gráfico dessa função está na Figura 2.

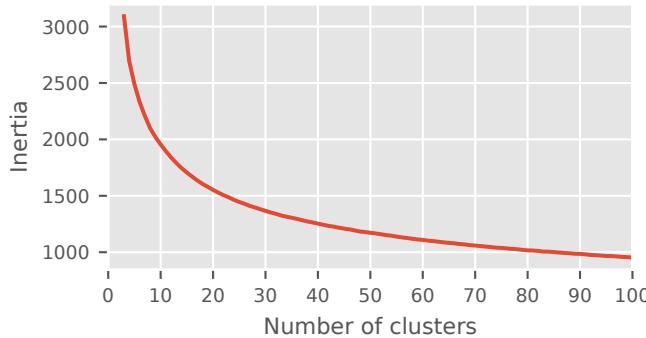


Figura 2. Valor da inércia por número de grupos

### B. Agrupamento

Com o número de grupos escolhido, o *k*-means foi aplicado aos dados. A distribuição de jogadores por grupo está bem equilibrada, como mostrado na Figura 3. Os centroides de cada grupo também foram analisados, com o intuito de verificar se os grupos são coesos. Estes representam perfis de jogadores completamente diferentes, como mostrado na Figura 4. Isso também mostra que os grupos encontrados pelo algoritmo são diferentes.

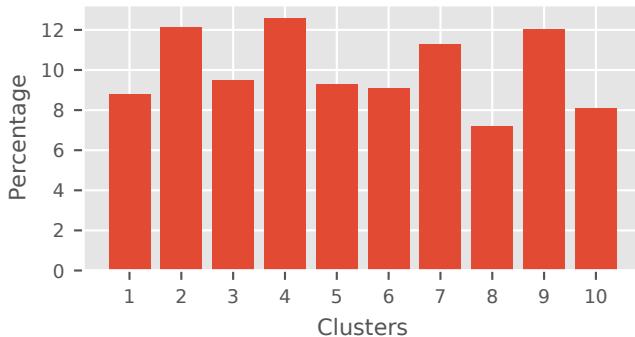


Figura 3. Distribuição de jogadores por grupo

Para avaliar a qualidade do agrupamento foi utilizado o coeficiente de silhueta [6]. Esse coeficiente varia entre  $[-1.0, 1.0]$ , e valores mais próximos de 1.0 indicam um bom agrupamento. A Figura 5 mostra a distribuição do coeficiente de silhueta dos jogadores por grupo. Os valores ficaram equilibrados para cada grupo, o que mostra a boa qualidade do agrupamento e corrobora o número de grupos escolhido para o *k*-means.

### C. KDA versus Clusters

As análises feitas foram com o intuito de verificar se realmente existiam grupos com perfis diferentes nos dados,

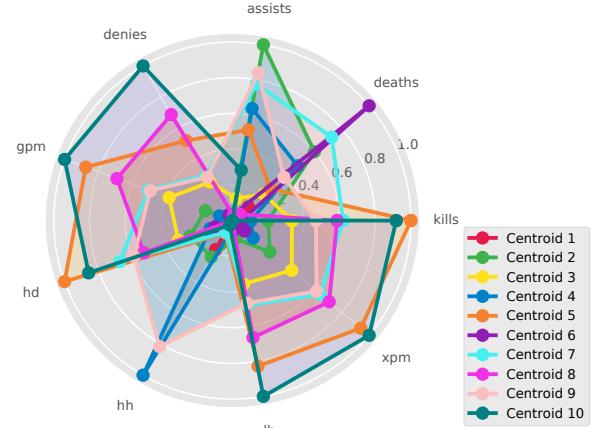


Figura 4. Atributos de cada centroide

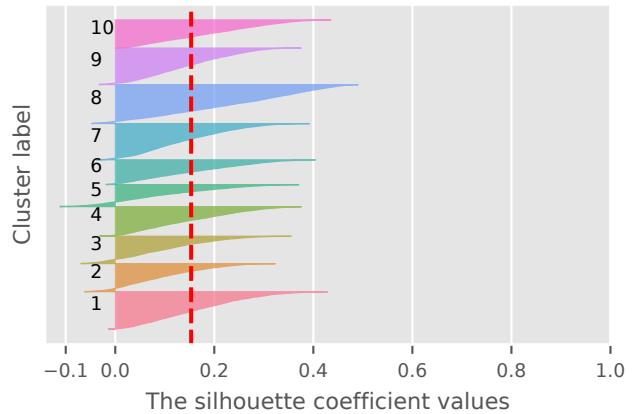


Figura 5. Distribuição do coeficiente de silhueta por grupo

o que se mostrou verdade. Então o KDA, que é a métrica mais utilizada no DOTA2, consegue capturar todas essas diferenças? Como mostrado na Figura 6, existem dois grupos com a CDF<sup>3</sup> quase idênticas. Isso mostra que o modo de avaliar jogadores utilizado atualmente não consegue diferenciar grupos distintos.

Isso sugere a criação de uma nova métrica, mais genérica e que consiga capturar as diferenças entre os grupos. Mas antes disso, é necessário fazer uma seleção dos atributos que irão compô-la. A técnica utilizada para isso e seus resultados serão discutidos na próxima seção.

## V. SELEÇÃO DE ATRIBUTOS

Antes de propor uma métrica, é necessário escolher os atributos que irão compô-la. Poderiam ser todos que existem na base, que são apenas nove, mas isso deixaria a métrica pouco intuitiva para o uso geral. Então, além dela ser robusta e conseguir capturar as diferenças entre perfis de jogadores, ela precisa ser compacta. Neste caso, essa seleção poderia

<sup>3</sup>Do inglês Cumulative Distribution Function

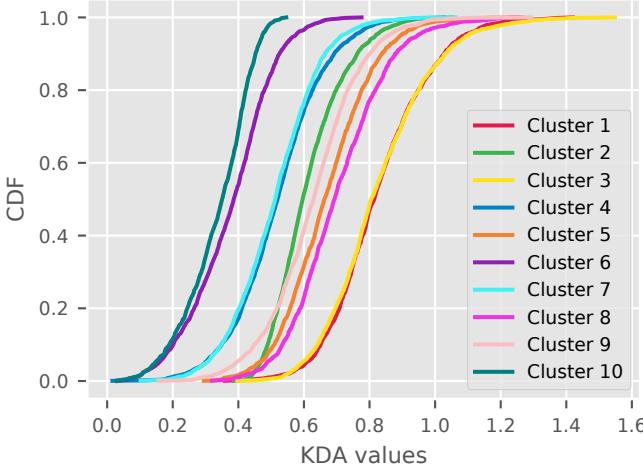


Figura 6. CDF da métrica KDA por grupo

ser por força bruta, mas os dados podem ser escalados no futuro, podendo conter muito mais informações e jogadores. Por esses motivos, foi proposta uma metodologia para a seleção de atributos que será explicitada nas seções subsequentes.

#### A. Justificativa / Escolha da técnica

Existem algumas heurísticas para a seleção de atributos em problemas de clusterização. Entre as encontradas na literatura, foi escolhido o GRASP [7] para fazer a busca da solução utilizando o coeficiente de silhueta como função de avaliação [8]. Essa escolha foi feita pela simplicidade do GRASP e da silhueta, além dos resultados dessa heurística serem satisfatórios.

#### B. Formulação do problema

O problema foi modelado como *Set Packing Problem*, e consiste em encontrar  $c \subset S$  que maximize a silhueta, onde  $S$  é o conjunto com todos atributos da base e  $tam\_min \leq |c| \leq tam\_max$ , onde  $tam\_min$  e  $tam\_max$  são parâmetros informados. Se dois atributos  $a_1, a_2 \in S$  possuem módulo da correlação acima de um limiar  $th$ , eles não podem aparecer juntos em  $c$ . Essa restrição foi feita por que, se  $a_1$  e  $a_2$  possuem alta correlação, eles representam informações muito parecidas, e isso causa redundância, o que não é desejado para a métrica. O valor utilizado para  $th$  foi 0.75, e foi encontrado empiricamente.

#### C. GRASP

O procedimento construtivo do GRASP proposto garante a geração somente de soluções viáveis, e quando se encontra uma solução inviável, esta é penalizada pelo algoritmo de acordo com o número de restrições violadas. Se as restrições relacionadas ao  $tam\_min$  e  $tam\_max$  forem violadas, aplica-se uma penalidade como se a solução violasse todos

os pares de restrições mais um, pois soluções que violam o tamanho não são desejadas, e devem ser evitadas ao máximo.

O custo  $C_{ins}(a)$  de inserção de  $a \in S$  em uma solução é dado por:

$$C_{ins}(a) = \frac{10 \cdot var(a)}{1 + NR(a)},$$

onde  $var(a)$  é a variância de  $a$  e  $NR(a)$  é o número de restrições em que  $a$  está envolvido.

O custo de uma solução  $c \in S$  é composto pela silhueta da clusterização utilizando o  $k$ -means diminuída de um fator de penalidade, e é dado por:

$$C_{sol}(c) = sil(k\text{-means}(c)) - pen(c),$$

sendo  $pen(c) = \log(1 + NV(c))$ , onde  $NV(c)$  o número de restrições violadas por  $c$ . Se  $c$  viola restrições de tamanho,  $NV(c) = \sum_{a \in c} NR(a) + 1$ . Note que, quando  $NV(c) = 0$ ,  $pen(c) = 0$ . Os algoritmos 2, 3, 4 e 5 mostram o GRASP, o procedimento construtivo, a busca local e a atualização da elite, respectivamente.

---

#### Algoritmo 2: GRASP para Seleção de Atributos (*Set Packing Problem*)

---

```

entrada:  $S$ : lista de itens (atributos)
entrada:  $R$ : matriz de restrições
entrada:  $tam\_min$ : tamanho mínimo para um conjunto de atributos viável
entrada:  $tam\_max$ : tamanho máximo para um conjunto de atributos viável
entrada:  $\alpha$ : fator de flexibilização da estratégia gulosa
entrada:  $t_e$ : tamanho do conjunto de elite
entrada:  $M$ : número máximo de iterações
entrada:  $m$ : número máximo de iterações sem melhorias na elite ou melhor solução encontrada
saída :  $e$ : conjunto de elite, com as  $t_e$  melhores soluções encontradas

Inicializa  $e$ ;
 $i \leftarrow 0$ ;
enquanto  $i < M$  faça
     $c \leftarrow$ 
        proc_construtivo( $S, R, tam\_min, tam\_max, \alpha$ );
         $e \leftarrow$  atualiza_elite( $c, e, t_e$ );
         $c \leftarrow$  busca_local( $c, S, tam\_min, tam\_max, m$ );
         $e \leftarrow$  atualiza_elite( $c, e, t_e$ );
    se  $m$  iterações sem melhorias então
        | Interrompe;
    fim
     $i \leftarrow i + 1$ ;
fim
retorna  $e$ 

```

---

---

**Algoritmo 3:** GRASP: Procedimento Construtivo

**entrada:**  $S$ : lista de todos os itens,  $R$ : matriz de restrições,  $tam\_min$  e  $tam\_max$ : tamanhos mínimo e máximo para  $c$  viável,  $\alpha$ : fator de flexibilização da estratégia gulosa

**saída :**  $c$ : solução candidata

```

 $LRC \leftarrow \alpha\%$  itens  $i$  de  $S$  com menor custo de inserção  $C_{ins}(i)$ ;
 $c \leftarrow \emptyset$ ;
enquanto  $LRC \neq \emptyset$  faz
     $el \leftarrow$  item retirado aleatoriamente de  $LRC$ ;
     $c \leftarrow c \cup \{el\}$ ;
    se  $tam(c) = tam\_max$  então
        | Interrompe laço;
    fim
    // O passo abaixo garante a
    // construção somente de soluções
    // viáveis
    Remove de  $LRC$  todos os itens ainda presentes
    que encontram-se em pares de  $R$  conjuntamente
    com  $el$ ;
     $LRC \leftarrow \alpha\%$  itens  $i$  de  $LRC$  com menor custo de
    inserção  $C_{ins}(i)$ ;
fim
se  $tam(c) < tam\_min$  então
    | Reiniciar processo todo;
fim
retorna  $c$ 
```

---

**D. Resultados da seleção**

As dez melhores soluções encontradas pelo GRASP estão na Tabela II. Dessas soluções, foram escolhidas as duas com o melhor fitness e a solução com os seguintes atributos: assists, deaths, denies, gpm, hh. Essa última escolha foi feita por que esses cinco atributos são os que mais aparecem nas melhores soluções. Os passos realizados para a elaboração das métricas serão detalhados na próxima seção.

Tabela II  
DEZ MELHORES SOLUÇÕES ENCONTRADAS PELO GRASP

Solução	Fitness
deaths,gpm,hh	0,189951
deaths,hh,xpm	0,187349
deaths,denies,hh,xpm	0,170163
assists,deaths,gpm,hh	0,150208
deaths,denies,gpm,hh	0,149013
assists,deaths,denies,gpm,hh	0,145158
assists,deaths,hh,lh	0,130658
deaths,denies,hh,kills	0,119486
assists,deaths,hh,xpm	0,119165
assists,deaths,denies,hh	0,113329

---

**Algoritmo 4:** GRASP: Busca Local

**entrada:**  $c$ : solução candidata,  $S$ : lista de todos os itens,  $tam\_min$  e  $tam\_max$ : tamanhos mínimo e máximo para  $c$  viável,  $m$ : número máximo de iterações sem melhorias

**saída :**  $c$ : solução candidata

```

 $i \leftarrow 0$ ;
enquanto  $i < m$  faz
     $viz \leftarrow$  cópia de  $c$ ;
    se  $tam(c) \leq tam\_min$  então
        | Acrescenta um elemento aleatoriamente em  $viz$ ;
    fim
    senão se  $tam(c) \geq tam\_max$  então
        | Retira um elemento aleatoriamente em  $viz$ ;
    fim
    senão se  $i$  está suficientemente longe de  $m$  então
        | Sorteia item  $el$  de  $S$  e retira  $el$  de  $viz$ , se
        estiver presente ou acrescenta  $el$  em  $viz$ , caso
        contrário;
    fim
    senão
        | Sorteia item  $ela$  de  $S$  e retira  $ela$  de  $viz$ , se
        estiver presente ou acrescenta  $ela$  em  $viz$ ,
        caso contrário;
        | Sorteia item  $elb$  de  $S$  e retira  $elb$  de  $viz$ , se
        estiver presente ou acrescenta  $elb$  em  $viz$ , caso
        contrário;
    fim
    se  $C_{sol}(viz) > C_{sol}(c)$  então
        |  $c \leftarrow viz$ ;
        |  $i \leftarrow 0$ ;
    fim
    senão
        |  $i \leftarrow i + 1$ ;
    fim
fim
retorna  $c$ 
```

---

**VI. MÉTRICA PROPOSTA**

Dado os melhores subconjuntos de atributos da base de dados, irá ser calculados o coeficiente de correlação  $\tau$  de Kendall entre cada um deles, que está entre  $[-1, 1]$ , onde pares com concordância perfeita possuem  $\tau = 1$  e os com discordância perfeita têm  $\tau = -1$ . Isso será feito para garantir que esses atributos não representam a mesma informação. Os resultados estão na Tabela III.

---

**Algoritmo 5:** GRASP: Atualização da Elite

---

**entrada:**  $c$ : solução candidata,  $e$ : conjunto elite,  $t_e$ : tamanho do conjunto elite  
**saída :**  $e$ : elite atualizada

```

se  $c \notin e$  então
    se  $\text{tam}(e) < t_e$  então
        | Acrescenta  $c$  em  $e$ ;
    fim
    senão se  $C_{\text{sol}}(c) > \min(\{C_{\text{sol}}(x) \mid \forall x \in e\})$ 
        então
            | Remove  $x$  com pior custo de  $e$ ;
            | Acrescenta  $c$  em  $e$ ;
    fim
fim
retorna  $e$ 
```

---

Tabela III  
KENDALL TAU ENTRE OS CONJUNTOS DE ATRIBUTOS

Conjunto A	Conjunto B	P-value	Kendall Tau
assists,deaths,denies, gpm,hh	assists,deaths,kills	0	0,56
assists,deaths,denies, gpm,hh	deaths,gpm,hh	0	0,67
assists,deaths,denies, gpm,hh	deaths,hh,xpm	0	0,63
assists,deaths,kills	deaths,gpm,hh	0	0,35
assists,deaths,kills	deaths,hh,xpm	0	0,35
deaths,gpm,hh	deaths,hh,xpm	0	0,87

Como o resultado da correlação entre os três conjuntos e o conjunto do KDA não foi alta, todos eles serão utilizados nas análises. O objetivo é propor métricas com eles e avaliações. Como apenas o atributo *deaths* é do tipo que precisa ser minimizado em uma partida, ou seja, quanto menos, melhor, ele será fator divisivo em todas as métricas e será acrescido de um, para não ter a possibilidade de ocorrer divisões por zero. Os outros entrarão somando no numerador da métrica, seguindo o padrão do KDA. As métricas propostas são:

$$ADG = \frac{\text{assists} + \text{denies} + \text{gpm} + \text{hh}}{1 + \text{deaths}},$$

$$G = \frac{\text{gpm} + \text{hh}}{1 + \text{deaths}},$$

$$X = \frac{\text{xpm} + \text{hh}}{1 + \text{deaths}}.$$

Na próxima seção, será detalhado o modo que foi utilizado para comparar essas métricas, e como foi feita a escolha da melhor delas.

## VII. RESULTADOS

A primeira análise feita nas métricas foi se elas realmente capturam perfis diferentes de jogadores em relação ao KDA. Como mostrado na Figura 7, os dez melhores jogadores em função do KDA possuem um perfil mais agressivo, i.e., que tenham os atributos *kills* e *hero damage* altos e *hero healing* baixo. Os dez melhores das métricas ADG, G e X estão nas Figuras 8, 9 e 10, respectivamente. Note que o ADG consegue capturar jogadores com diferentes perfis, ou seja, com *hero healing* mais alto, mas sem perder a qualidade nos outros atributos. Isso a torna uma métrica mais equilibrada se comparada com o KDA.

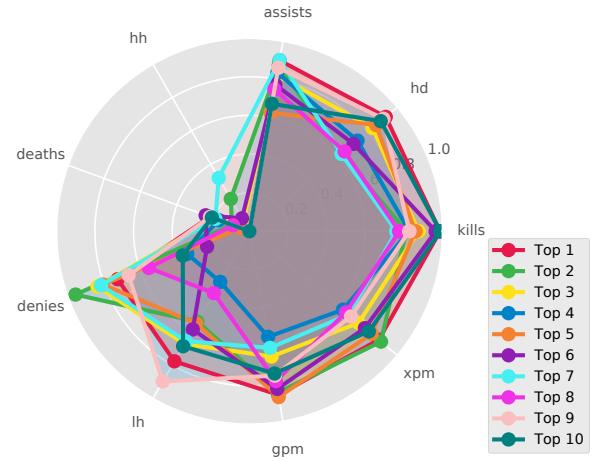


Figura 7. Dez melhores jogadores em relação à métrica KDA

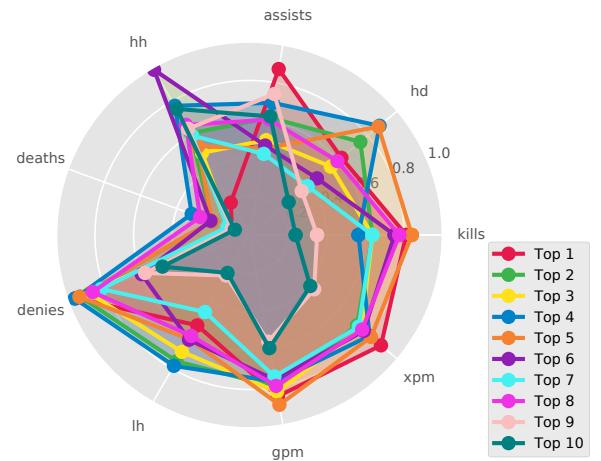


Figura 8. Dez melhores jogadores em relação à métrica ADG

Em relação a G e X, elas possuem um comportamento muito parecido, com um *hero healing* muito alto, e com os outros atributos menores, exceto por um jogador com o atributo *assists* elevado em X. Isso é esperado, já que a correlação entre os conjuntos de atributos que formam as duas métricas é de 0,87, como mostrado da Tabela

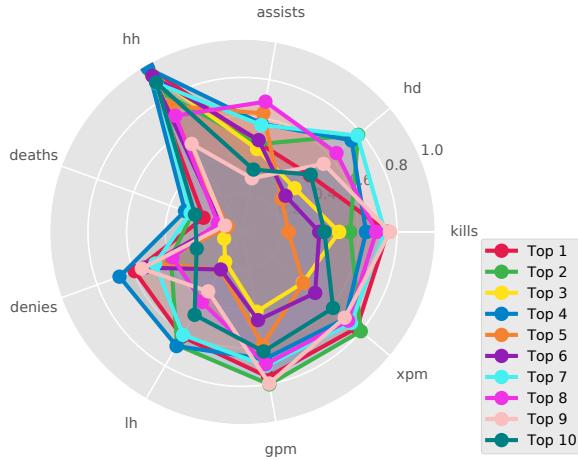


Figura 9. Dez melhores jogadores em relação à métrica G

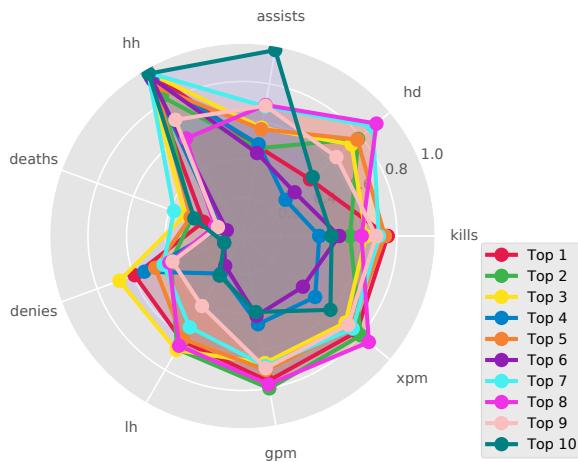


Figura 10. Dez melhores jogadores em relação à métrica X

III. As duas métricas capturam perfis de jogadores quase complementares ao KDA, o que também é esperado, já que a correlação entre cada uma das duas e o conjunto que forma o KDA é de 0,35, como mostrado na Tabela III.

Como o agrupamento gera grupos distintos (ver seção IV), uma métrica boa deve gerar CDFs distintas para cada grupo, o que não acontece no KDA, como mostrado na Figura 6. As CDFs do ADG, G e X estão nas Figuras 11, 12 e 13, respectivamente. De todas as métricas, a única que cumpriu o objetivo proposto nessa parte foi a ADG. Além dela capturar perfis de jogadores mais equilibrados, ela consegue diferenciar melhor os grupos encontrados pelo  $k$ -means. Por esses motivos, essa será a métrica proposta por esse trabalho.

### VIII. CONCLUSÕES

Neste trabalho, estudou-se o problema de avaliação de jogadores de DOTA2, e como criar uma métrica que consegue capturar os vários perfis existentes de jogadores.

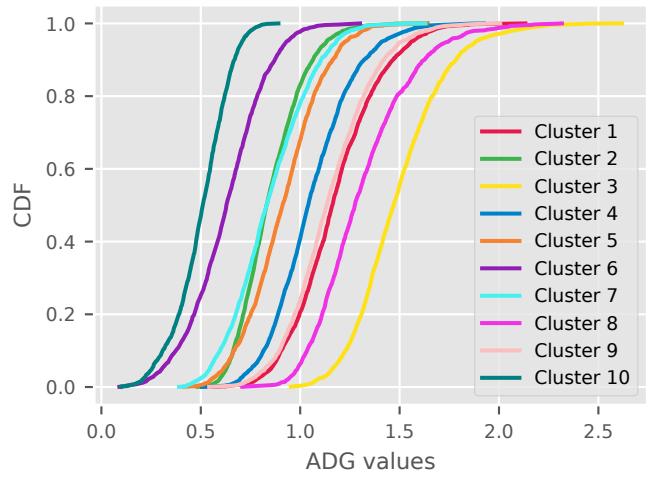


Figura 11. CDF da métrica ADG por grupo

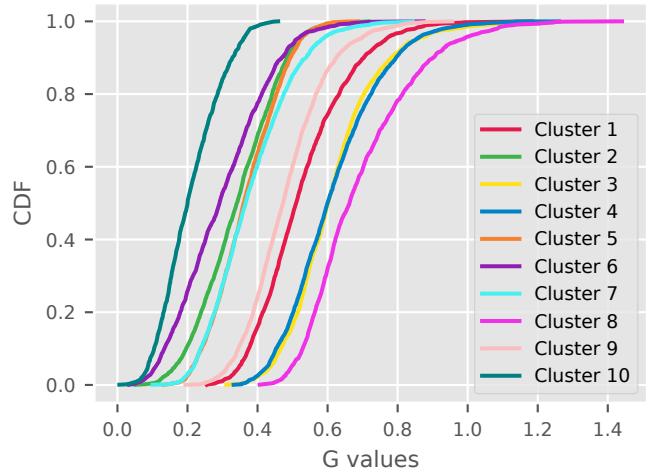


Figura 12. CDF da métrica G por grupo

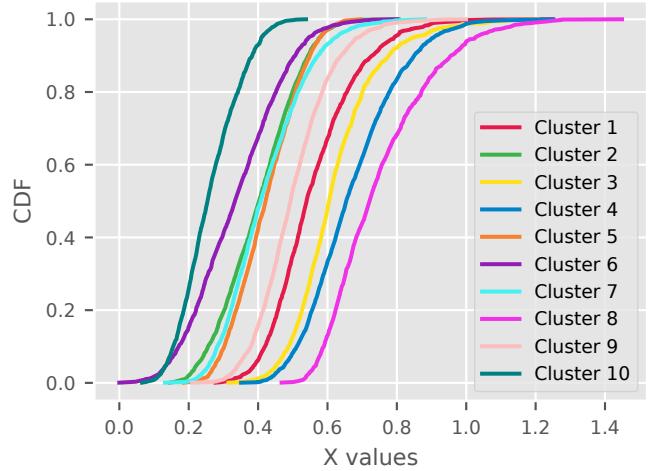


Figura 13. CDF da métrica X por grupo

Para resolvê-lo, foi utilizada uma técnica de aprendizagem não-supervisionada, mas especificamente o  $k$ -means, para descobrir se existem grupos coesos nos dados. Após isso, foi feita uma seleção de atributos utilizando GRASP para achar o melhor subconjunto de atributos que iriam fazer parte da métrica proposta. Foram escolhidos os dois conjuntos de atributos com o melhor *fitness* e o que tinham os melhores cinco atributos. Esses conjuntos não têm o coeficiente  $\tau$  de Kendall alto com o KDA, logo, eles representam informações diferentes. Esses atributos derivaram três métricas, chamadas ADG, G e X.

Para encontrar a melhor métrica entre as três, foram utilizadas duas abordagens: a análise dos dez melhores jogadores por métrica e a análise da CDF das métricas por grupo. Foi mostrado que a métrica ADG consegue capturar melhor os diferentes perfis de jogadores se comparada com as outras métricas.

É importante salientar que este é um trabalho em andamento. Atualmente, encontra-se em fase de desenvolvimento um sistema para a avaliação da métrica proposta em jogadores profissionais de DOTA2. Feito isso, este trabalho será bastante relevante para a comunidade de DOTA2, que poderá utilizar uma métrica melhor embasada e mais justa em suas análises.

Além disso, este trabalho abre oportunidades de trabalhos futuros, como por exemplo, criar um sistema para prever resultados de partidas de DOTA2, utilizando essa métrica como função de avaliação dos jogadores. Também pode ser proposto um novo sistema para ranqueamento de jogadores de DOTA2 utilizando o ADG.

#### AGRADECIMENTOS

Um agradecimento especial ao Felipe Dias por toda a ajuda com assuntos relacionados a DOTA2 e por várias ideias que inspiraram esse trabalho.

#### REFERÊNCIAS

- [1] PCGamer. (2017) Esports was worth 1.5 billion in 2017. [Online]. Available: <https://www.pcgamer.com/esports-was-worth-15-billion-in-2017/>
- [2] E.-S. Earnings. (2018) Top games awarding prize money. [Online]. Available: <https://www.esportsearnings.com/games>
- [3] D. P. P. Tracker. (2018) Dota2 prize pool tracker. [Online]. Available: <http://dota2.prizetrac.kr/international2017>
- [4] J. MacQueen *et al.*, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.
- [5] M. J. Zaki and J. Wagner Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, May 2014.
- [6] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.
- [7] T. A. Feo and M. G. Resende, “Greedy randomized adaptive search procedures,” *Journal of global optimization*, vol. 6, no. 2, pp. 109–133, 1995.
- [8] J. R. Marino and L. H. Lelis, “Feature selection as state-space search: An empirical study in clustering problems,” in *Eighth Annual Symposium on Combinatorial Search*, 2015.