

Estudo de Caso: Análise do Método C4.5 na Predição do Papel de Jogadores de *League of Legends*

Guilherme F. Gaiardo¹, Luísa P. Lucca¹, Brenda S. Santana¹,
Maurício M. Donato¹, Vinícius F. Garcia¹, Ana T. Winck²

¹Departamento de Computação Aplicada -
Universidade Federal de Santa Maria (UFSM)

²Departamento de Ciências Exatas e Sociais Aplicadas -
Fundação Universidade Federal de Ciências da Saúde de Porto Alegre

{ggaiardo, llucca, bsantana, mdonato, vfulber, ana}@inf.ufsm.br

Abstract. *Multiplayer Online Battle Arena (MOBA) games have been in increasing popularity since the middle of the year of 2005. In this context, League of Legends (LoL) has become an important exponent in the MOBAs' universe. This paper aims to analyze the possibility of detecting the role assumed by players during the matches, for this purpose it's used statistics and characteristics from best players by the official ranking. From the execution of several configurations of C4.5 algorithm we try to detect gaming patterns and understand the generated results.*

Resumo. *Jogos do gênero Multiplayer Online Battle Arena (MOBA) apresentam-se em crescente popularização desde meados do ano de 2005. O jogo intitulado League of Legends (LoL) se tornou um importante expoente no universo dos MOBAs. Este artigo objetiva analisar a possibilidade de detecção do papel desempenhado pelos jogadores durante as partidas, para isso utiliza-se estatísticas e características de jogadores destaques pelo ranque oficial. A partir da execução de diversas configurações do algoritmo C4.5 busca-se detectar padrões de jogo e compreender os resultados gerados.*

1. Introdução

Há aproximadamente duas décadas, o gênero de jogos *Multiplayer Online Battle Arena* (MOBA) surgia através do título "Herzog Zwei". Desde então, diversas adaptações moldaram o estilo e jogabilidade dos MOBAs, sendo extremamente competitivo e estratégico. Títulos do gênero, aliados aos FPS, dominam o mundo do chamado Esporte Eletrônico ou *e-sport*, movimentando um novo mercado com economia diferenciada [Seo 2013].

O MOBA intitulado *League of Legends* (LoL) foi desenvolvido pela Riot Games e lançado em Outubro de 2009. Desde então popularizou-se e em 2012 alcançou a marca de jogo para computadores pessoais mais jogado no mundo [Gaudiosi 2012]. Em 2011 foi realizado o primeiro torneio mundial de LoL, o mesmo se repete anualmente desde então.

O jogador de *League of Legends* assume um papel específico dentro da arena durante uma partida. São cinco os papéis possíveis: *Top*, *Mid*, *AD Carry*, *Support* e *Jungle*. Cada papel desenvolve funções principais únicas e outras funções secundárias comuns durante o tempo de jogo.

O objetivo deste trabalho é, a partir de informações estatísticas de jogadores de *League of Legends*, determinar o papel que é desenvolvido pelos mesmos durante uma partida. O método C4.5 [Quinlan 1993], na figura da implementação intitulada J48 [Bhargava et al. 2013], implementado na ferramenta Weka [Holmes et al. 1994] foi utilizado para a geração das árvores de decisão.

Este trabalho é dividido em seis seções onde a Seção 2 apresenta os trabalhos relacionados. A Seção 3 traz informações sobre a base de dados utilizada. A seção 4 exhibe os testes executados e os resultados obtidos. A Seção 5 exhibe uma discussão geral sobre os resultados. Por fim, a seção 6 conclui o artigo.

2. Trabalhos Relacionados

Mineração de dados trouxe técnicas que já são amplamente utilizadas no contexto esportivo desde a realização de *scouting* até predição de resultados e medição de desempenho de equipes e jogadores [Solieman 2006]. Todo esse processo consiste no gerenciamento do conhecimento esportivo e traz inúmeros benefícios na comparação de resultados de equipes que aplicam essas técnicas em relação aos que não aplicam [Schumaker 2010].

Com o crescimento e popularização do *eSport* nasce a área hoje chamada de *eSports analytics* objetivando analisar dados de diferentes partidas através de técnicas como mineração de dados com objetivos como aprimoramento de desempenho e definição de estratégias [Schubert et al. 2016].

Vários trabalhos utilizam o gênero MOBA como material para aplicação de técnicas de *eSports analytics*. Em [Van De Bovenkamp et al. 2013] os autores analisam dados de jogos que apresentam como característica o *Online Social Gaming*, onde a interação social afeta diretamente a experiência de jogo, afim de extrair gráficos de interação e aprimorar a sistemas de combinação de jogadores.

Em [Almeida et al. 2017] é definido um processo de descoberta de conhecimento para prever a chance de uma equipe de jogadores ganhar baseado em padrões estratégicos do jogo Defense of the Ancients (DOTA) 2, também do gênero MOBA. Os autores deixam claro que as mesmas técnicas apresentadas podem ser aplicadas a demais jogos do gênero MOBA, como *League of Legends* e *Heroes of Newerth*.

3. Base de Dados

A base de dados foi obtida através de um *script* programado em linguagem Python utilizando a Interface de Programação de Aplicação (API) fornecida pela RIOT Games⁶. Em um primeiro momento foram obtidos todos os jogadores ranqueados como *Master* e *Challenger*, por serem considerados os melhores de cada região e, portanto, terem um estilo de jogo bem definido.

A partir da lista dos melhores jogadores das regiões do Brasil e América do Norte foram recuperados os dados das partidas ranqueadas de cada jogador da lista. A lista das partidas inclui estatísticas de todos os campeões separadamente e a estatística agregada de todos os personagens jogáveis. Os atributos selecionados foram resultado das estatísticas agregadas consideradas mais relevantes para o cenário competitivo.

A base de dados conta com 1711 registros e 12 atributos em cada registro que

⁶ <https://github.com/GuiGaiardo/BDLLOL>

identificam o estilo do jogador analisado. Essa base de dados apresenta originalmente 11 atributos contínuos e 1 atributo discreto. Todos os atributos contínuos representam estatísticas obtidas através da análise das partidas ranqueadas que o jogador participou. Os atributos selecionados são descritos a seguir:

- **ROLE:** papel desempenhado pelo jogador, atributo discreto utilizado como objetivo da indução.
- **KPS:** média de personagens mortos pelo jogador por partida.
- **DPS:** média de mortes do jogador por partida.
- **APS:** média de assistências realizadas pelo jogador por partida.
- **MKPS:** média de personagens não jogáveis mortos pelo jogador por partida.
- **NMKPS:** média de personagens não jogáveis neutros mortos pelo jogador por partida.
- **DDPS:** média de dano causado pelo jogador por partida.
- **DTPS:** média de dano recebido pelo jogador por partida.
- **GEPS:** média de ouro recuperado pelo jogador por partida.
- **HPS:** média de curas recebidas pelo jogador por partida.
- **MDDPS:** média de dano mágico causado pelo jogador por partida.
- **PDDPS:** média de dano físico causado pelo jogador por partida.

Para a realização dos testes, duas versões da base de dados foram utilizadas: a primeira versão, nomeada DB1, utiliza os dados integralmente como foram obtidos, já a segunda versão, ou DB2, discretiza todos os atributos através do discretizador padrão da ferramenta Weka.

4. Testes e Resultados

Os testes foram executados utilizando a implementação do método C4.5 [Quinlan 1993] chamado J48 [Bhargava et al. 2013], este algoritmo é disponibilizado nativamente pela ferramenta Weka. O Fator de Confiança (FC) sofreu variação entre 0,05 e 0,25. Fatores de confiança superiores a 0,25 não foram utilizados afim de manter a capacidade de generalização da árvore. Para o DB2, tanto múltiplos cortes quanto cortes binários foram considerados nos testes.

As métricas selecionadas para a avaliação dos resultados consistem de acurácia, precisão e área ROC. Todos os resultados foram obtidos através de validação cruzada com 10 *folds* através da ferramenta Weka.

As bases de dados foram inicialmente analisadas utilizando todos os atributos disponíveis. As Figuras 1 e 2 apresentam os melhores resultados obtidos para este cenário onde cortes binários foram utilizados no DB2.

Apesar da aparente semelhança entre as execuções com fator de confiança de 0,05 e 0,1, o DB1 apresenta melhores resultados em todos os aspectos avaliados quando esse fator é configurado em 0,05. O DB2, ao contrário, conta com um acréscimo nos valores de acurácia e precisão quando utilizado 0,1 como fator de confiança, entretanto uma ligeira queda na área da curva ROC é verificada.

A partir da análise da matriz de confusão gerada pelas execuções é possível determinar os pontos com maior ocorrência de erros. A confusão entre os papéis *Mid*, *Top* e *AD Carry* representam 70% do número absoluto de registros classificados incorretamente para o DB1 (0,05 FC) e 68% para o DB2 (0,1 FC). Outro problema detectado foi na interpretabilidade do modelo induzido, as árvores geradas são

demasiadamente grandes.

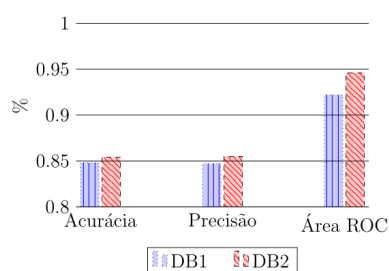


Figura 1: Resultados J48 (FC 0.05)

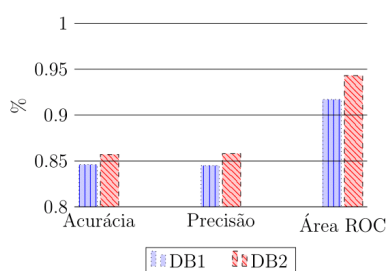


Figura 2: Resultados J48 (FC 0.10)

4.1. Filtragem de Atributos

A filtragem de atributos mostra-se como uma opção para a criação de uma árvore diminuta, com menor complexidade e maior interpretabilidade. Também busca-se maior acurácia e precisão ao eliminar atributos ruidosos ou que geram pouco ou nenhum ganho de informação.

Duas estratégias principais foram adotadas: a análise do ranqueamento dos atributos, utilizando dados recuperados de vários avaliadores disponíveis no Weka, e a análise da correlação e redundância das informações utilizando o avaliador intitulado *Cfs Subset*, nesse caso os métodos de busca de atributos foram *Greedy Stepwise* e *Best First*. Toda a etapa de filtragem de atributos foi realizada na ferramenta Weka.

O ranqueamento foi realizado analisando dados relacionados ao atributo alvo, de correlação, de proporção de ganho, ganho de informação, classificador OneR, distância de valores e incerteza simétrica. O resultado da avaliação, em ordem decrescente de relevância dos atributos, é apresentado na Tabela 1.

DB1	MKPS, NMKPS, DDPS, APS, PDDPS, DTSP, KPS, GEPS, HPS, MDDPS, DPS
DB2	MKPS, NMKPS, DDPS, APS, PDDPS, DTSP, KPS, GEPS, HPS, MDDPS, DPS

Tabela 1. Atributos Selecionados Via Ranqueamentos

Testes reduzindo gradativamente o número de atributos foram realizados. De forma geral, os resultados obtidos pelos processos de ranqueamento se confirmaram em relação a ordem de relevância dos atributos, exceto para o atributo MDDPS, que se mostrou essencial para a indução de um bom modelo.

As Figuras 3 e 4 apresentam os melhores resultados obtidos através da remoção de atributos norteadas pelos algoritmos utilizados para ranqueamento. Para o DB1 foram removidos os atributos DPS, GEPS, HPS, KPS, PDPS e APS. Para o DB2 foram removidos os atributos DPS e HPS.

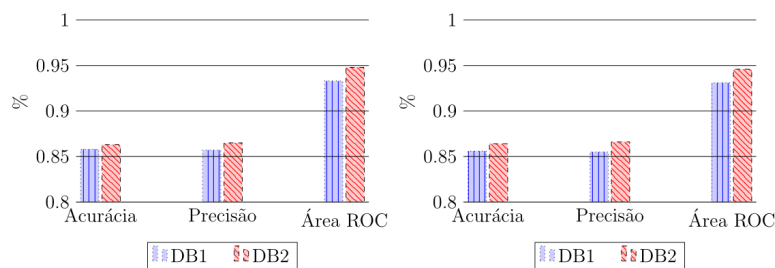


Figura 3: Ranker J48 (FC 0.05)

Figura 4: Ranker J48 (FC 0.10)

O avaliador *Cfs Subset* aliado aos métodos de busca *Best First* e *Greedy Stepwise* também foram aplicados nas bases de dados. O método *Best First* foi executado nos três modos possíveis: *forward*, *backward* e *bi-directional*. O método *Greedy Stepwise* foi configurado com limite de corte em 0,25. Os atributos selecionados para ambos os métodos são apresentados na Tabela 2. Os mesmos atributos foram escolhidos tanto para o DB1 quanto para o DB2.

Best First	Forward	KPS,APS,MKPS,NMKPS,DDPS,DTPS,HPS,MDDPS,PDDPS
	Backward	KPS,APS,MKPS, NMKPS, DTPS, HPS, MDDPS, PDDPS
	Bi-directional	KPS, APS,MKPS,NMKPS,DTPS,HPS,MDDPS,PDDPS
Greedy Stepwise	Threshold = 0,25	KPS,APS,MKPS,NMKPS,DDPS,DTPS,HPS,MDDPS,PDDPS

Tabela 2. Atributos Selecionados Via *Cfs Subset*

Os atributos selecionados com o método *Best First backward* e *bi-directional*, assim como com a configuração *forward* e o método *Greedy Stepwise* são os mesmos. Os resultados dos dois conjuntos obtidos são apresentados nas Figuras 5 e 6 onde utilizou-se como fator de confiança 0,1 e 0,25 respectivamente para DB1 e DB2.

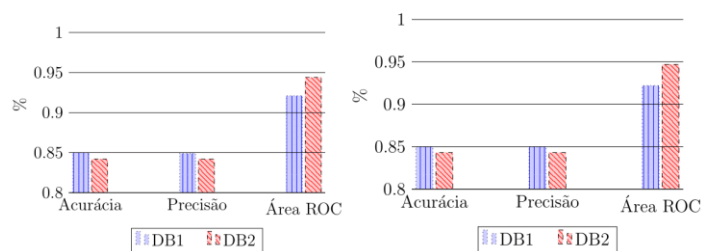


Figura 5: Backward e Bi-directional

Figura 6: Forward e Greedy Stepwise

4.2. Aplicação de Wrappers

A última técnica de seleção de atributos utilizada foi um *Wrapper* que testa diversos sub-conjuntos de atributos para encontrar o sub-conjunto gerando a melhor árvore com o algoritmo J48. Os parâmetros do J48 utilizados foram: fator de confiança 0.05 sem cortes binários. Os algoritmos de busca utilizados foram o *Greedy Stepwise* padrão e *Best First* com limite de 5 expansões sem melhora.

A Tabela 3 apresenta os resultados para o DB1. Ambos os algoritmos encontraram o mesmo sub-conjunto de atributos em todas as rodadas. Em comparação com o conjunto de atributos escolhido pela busca total, foram retirados os atributos DPS, GEPS e PDDPS, e adicionado o atributo HPS.

Best First e Greedy-Stepwise	MKPS,NMKPS,DTPS,HPS,MDDPS
Exhaustive Search	DPS,MKPS,NMKPS,DDPS,DTPS,GEPS,MDDPS,PDDPS

Tabela 3. Atributos Selecionados Via Wrapper para o DB1

A Figura 7 apresenta os resultados do J48 para o DB1 com os atributos selecionados pelas buscas rápidas e pela busca total. O resultado indica que a avaliação feita pelo algoritmo de busca acaba por supervalorizar atributos que diminuem o poder de generalização da árvore induzida, conforme indicam os resultados superiores da árvore com menor número de atributos.

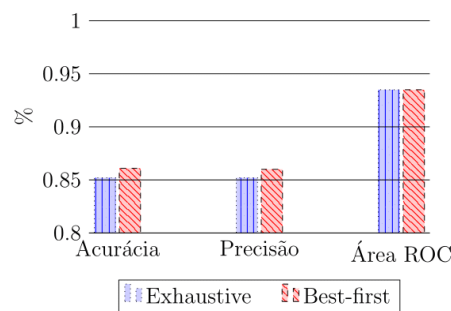


Figura 7: DB1

A Tabela 4 apresenta os resultados do Wrapper para o DB2. Os atributos selecionados com *Best First bi-directional* e *backward* foram os mesmos, também são expostos os atributos selecionados pelo método *Exhaustive Search*. Tanto o *Best First bi-directional* quanto a *Exhaustive Search* encontraram o mesmo conjunto de atributos ótimos.

Best First	Forward	KPS,DPS,MKPS,NMKPS,DDPS,DTPS,MDDPS,PDDPS
	Bi-directional	KPS,DPS,NMKPS,DDPS,DTPS,HPS,MDDPS,PDDPS
Greedy Stepwise	Threshold = 0,25	KPS,MKPS,NMKPS,DTPS,MDDPS,PDDPS
Exhaustive Search		KPS,DPS,NMKPS,DDPS,DTPS,HPS,MDDPS,PDDPS

Tabela 4. Atributos Selecionados Via Wrapper para o DB2

A Figura 8 mostra os resultados do J48 para o DB2 com os atributos selecionados pelas buscas rápidas e pela *Exhaustive Search*. Os resultados dos 3 conjuntos de atributos foram semelhantes, com um pequeno aumento na acurácia e precisão do *Best First forward*.

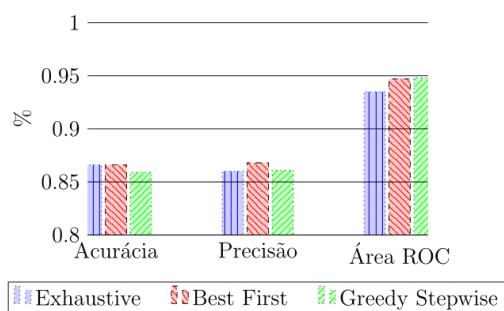


Figura 8: DB2

5. Discussão

Alguns atributos selecionados influenciavam muito pouco na predição das árvores de decisão, outros levavam a erros quando utilizados. Dessa forma o algoritmo J48 apresentou melhores taxas de acurácia, precisão e curva ROC quando aplicado filtros ou *wrappers* para realizar seleção de atributos nas bases de dados.

A análise da acurácia leva a determinar que o *Wrapper* utilizando as técnicas de *Exhaustive Search* ou *First* apresentou os melhores resultados, chegando a um valor de 0,866 para o DB2. Entretanto, tanto precisão, quanto curva ROC foram superiores quanto utilizado a técnica de *Best First* determinando uma melhor capacidade de generalização.

Em relação a aplicação de filtros, os resultados obtidos foram considerados satisfatórios. Muito próximo as taxas obtidas com *Wrappers*. Para o DB2, ao utilizar vários processos de ranqueamento para determinação dos atributos, obteve-se acurácia de 0,864.

Com a aplicação de filtros no DB1 não foi possível obter taxas de acurácia iguais ao DB2. Porém a utilização de atributos contínuos permitiu a retirada de vários atributos sem degradação da acurácia, precisão e curva ROC, gerando um modelo mais interpretável.

6. Conclusão

Com a popularização de jogos e esportes eletrônicos novas áreas de pesquisa que objetivam analisar e otimizar processos envolvidos na preparação, execução de tarefas e melhoramento de desempenho desses novos atletas se mostram cada vez mais necessárias.

Métodos de mineração de dados são adequados para a realização de análises e predições baseados em bases de dados que contenham informações específicas do nicho que se deseja atingir. Jogos do tipo MOBA são capazes de gerar diversos tipos de estatísticas e dados passíveis de passar por processos de mineração.

Este artigo realiza um estudo de caso que objetiva analisar a aplicabilidade do método C4.5 na predição de papéis desempenhados por jogadores do MOBA *League of Legends*. Para isso, diversas configurações de execução e algoritmos de filtragem foram aplicadas em uma base de dados inicial recuperada através de dados fornecidas pela própria desenvolvedora do jogo.

O método C4.5, na figura da implementação J48, se mostrou eficaz na predição

dos papéis desempenhados pelos jogadores dado diferentes conjuntos de atributos que caracterizam os mesmos. Em especial, a utilização de dados discretos obteve maior acurácia, porém, ao considerar atributos contínuos os modelos gerados adquiriram maior interpretabilidade.

Em trabalhos futuros objetiva-se executar análises estatísticas sobre os dados coletados. Também, a atualização da base de dados é necessária visto que os jogadores ranqueados como *Challenger* e *Master* podem mudar. Além disso, o chamado meta jogo, que define estratégias e ações dos jogadores durante a partida, também é mutável, sendo possível que variações nos atributos avaliados ocorram em análises futuras.

Referências

- Almeida, C. E., Correia, R. C., Eler, D. M., Olivete-Jr, C. Garci, R. E., Scabora, L. C., and Spandon, G. (2017). "Prediction of winners in moba games", In: Information Systems and Technologies (CISTI). 2017 12th Iberian Conference on, pages 1-6. IEEE.
- Bhargava, N., Sharma, G., Bhargava, R., and Mathuria, M. (2013). "Decision tree analysis on j48 algorithm for data mining.", Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering. 3(6).
- Gaudiosi, J. (2012). "Riot Games' league of legends officially becomes most played pc game in the world", Forbes, Jul, 11:2011.
- Holmes, G., Donkin, A., and Witten, I.H (1994). "Weka: a machine learning worckbench." In Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on, pages 357-361.
- Quinlan, J. R. (1993). "*C4.5: Programs for Machine Learning*". Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Schubert, M., Drachen, A., and Mahlmann, T. (2016). Esports analytics through encounter detection other sports.
- Schumaker, R. P., Solieman, O. K., and Chen, H. (2010). Sports knowledge management and data mining. *Annual review of information science and technology*, 44(1):115-157.
- Seo, Y. (2013). "Eletronic sports: A new marketing landscape of the experience economy". *Journal of Marketing Management*, 29(13-14):1542-1560.
- Solieman, O. K. (2006). "Data Mining in sports: A research overview". *Dept. of Management Information Systems*.
- Van De Bovenkamp, R., Shen, S., Iosup, A., and Kuipers, F. (2013). "Understanding and recommending play relationships in online social gaming". In *Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on*, pages 1-10, IEEE.