

ANÁLISE DE TÉCNICAS DE APRENDIZADO SUPERVISIONADO APLICADAS NA AVALIAÇÃO DE SUCESSO DE AÇÕES NO FUTEBOL DE ROBÔS

Bruno Bollos Correa¹, Flavio Tonidandel²

¹ Engenharia Elétrica, Centro Universitário FEI

² Ciência da Computação, Centro Universitário FEI
{uniebruncorrea, flaviot}@fei.edu.br

Resumo: Este projeto tem como objetivo analisar diferentes tipos de aprendizado de máquina supervisionado de regressão aplicados na decisão de ações do robô com o domínio da bola em uma partida de futebol de robôs da Small Size League (SSL). O conjunto de dados utilizado foi adquirido a partir de análises de logs de partidas da liga com o software LogAnalyzerRoboFEI-SSL, programa que foi desenvolvido para gerar informações de todos os passes e chutes a gol ocorridos durante uma partida. Os métodos: Regressão Linear, k-Vizinhos Mais Próximos (kNN), *Weighted-kNN* (w-kNN), Floresta Aleatória, *AdaBoost* e *Gradient Boosting* foram ajustados e testados. *Gradient Boosting* e w-kNN geraram os melhores modelos estimadores da jogada de chute; enquanto todos os modelos estimadores da jogada de passe obtiveram resultados similares nas predições das respostas. Devido a erros durante o processo de aquisição de dados e à incerteza do fenômeno, os modelos ficaram menos precisos; mesmo assim, os estimadores conseguiram detectar certos padrões e foram úteis a ponto de serem integrados no software de estratégia da equipe para auxílio na decisão de ações dos robôs durante as partidas.

1. Introdução

A categoria *Small Size League* do futebol de robôs da *RoboCup* é uma das categorias mais dinâmicas de toda a competição, isto ocorre pela descomplicada movimentação dos robôs aliada à velocidade dos mesmos e pelos passes e chutes rápidos e espontâneos. Por essa dinâmica de jogo, torna-se necessária a implementação de uma estratégia que tome decisões rápidas e inteligentes durante as partidas.

A solução do problema de tomada de decisões pode ser solucionada utilizando a abordagem de Aprendizado de Máquina, isto é, fornecendo um conjunto de exemplos de decisões, com seus atributos relevantes e respostas, ao programa de computador que poderá, então, utilizar essas informações para realizar novas predições no futuro.

Este estudo tem como objetivo desenvolver, analisar e implementar diferentes métodos de aprendizado de máquina supervisionado aplicados na avaliação do sucesso de ações de passe e chute a gol em partidas de futebol de robôs da categoria *Small Size League* da *RoboCup*.

2. Aprendizado de Máquina

Aprendizado de máquina (*Machine Learning* ou ML) são métodos computacionais que usam a experiência para melhorar a performance de algum processo, para achar

padrões escondidos ou para fazer previsões mais precisas sobre determinado fenômeno. Problemas de ML podem ser resolvidos reunindo um conjunto de dados de exemplos do fenômeno e, em seguida, construindo um modelo estatístico com base nesse conjunto de dados [1].

Existem três principais tipos de aprendizado de máquina: o aprendizado supervisionado, não supervisionado e por reforço [1]. Com o objetivo de avaliar jogadas em partidas de futebol, o aprendizado de máquina supervisionado é o mais indicado dentre os três.

Em um problema de aprendizado supervisionado, o conjunto de dados contém uma quantidade de amostras, com cada amostra composta de alguns atributos como entrada e uma resposta, e o objetivo é construir um modelo que possa estimar respostas de entradas genéricas condizentes com o conjunto de dados.

Para este estudo, o aprendizado supervisionado é abordado em um problema de regressão, o que significa que o modelo estima valores numéricos contínuos em suas respostas, em vez de respostas categorizadas, como seria em um problema de classificação [2].

Os métodos escolhidos para este estudo, a Regressão Linear é o único método linear, onde se busca encontrar a tendência linear entre os atributos e a resposta; dois deles (kNN e w-kNN) se baseiam em amostras do conjunto de dados para realizar previsões; os três últimos (Floresta Aleatória, *AdaBoost* e *Gradient Boosting*) são métodos de *Ensemble Learning*, onde se tem a agregação de vários estimadores fracos para se obter um estimador mais forte [1-4, 6].

3. Metodologia

A metodologia deste trabalho é separada em quatro etapas.

1. O primeiro passo foi a criação do conjunto de dados. Isso foi feito pela análise de vários arquivos de log de partidas da liga pelo *software LogAnalyzer*, programa desenvolvido para lidar com arquivos log de partidas e realizar análises. Como resultado, se obteve um conjunto de dados extenso com várias amostras de jogadas de passe e chute a gol.
2. Seguidamente, se estudou o conjunto de dados gerado com o intuito de remover amostras que apresentassem muitos ruídos ou incoerências.
3. Neste ponto se tinha dois conjuntos de dados, um para jogadas de chute a gol e outro para jogadas de passe. O passo seguinte foi estudar, criar e ajustar os modelos de ML com os 6 métodos escolhidos para este trabalho. Para esta

tarefa, se utilizou a biblioteca *scikit-learn* em Python.

- Após o treino e ajuste dos modelos de ML, estes foram implementados no software de estratégia da equipe para auxiliar na decisão de ações do robô durante a partida.

Para a aquisição de dados, o software *LogAnalyzer* detectava a ocorrência de jogadas de passe e chute, guardava atributos referentes às condições iniciais da jogada e, uma vez que a jogada era finalizada, a avaliava de acordo com seu desfecho, dando-lhe uma nota entre 0 e 250

Os atributos, em sua maioria, são medidas de ângulo e distância que outras equipes utilizam no processo de decisão de jogadas em suas estratégias. Para a jogada de chute, os 3 atributos adotados foram: o ângulo livre e distância da bola ao gol e a marcação do adversário à bola. Para a jogada de passe, os 8 atributos adotados foram: ângulo livre e distância da bola ao receptor e do receptor ao gol, marcação à bola, marcação ao receptor, ângulo para chute de primeira e variação da bola no eixo x .

Durante o ajuste dos modelos, o conjunto de dados foi subdividido em dois: o de treino, utilizado como base para a criação dos modelos; e o de teste, utilizado para validação do modelo. A qualidade de um modelo foi dada em função do coeficiente de determinação (R^2) no conjunto de dados de teste. O R^2 é uma métrica que computa o quão bem um modelo regressor está ajustado e a fração de variação da resposta justificada pelo mesmo em uma escala entre 0 e 1 [5].

Outros parâmetros analisados nos modelos incluem o tempo médio de previsão de 10 mil entradas aleatórias (t_{pred}) e a importância de cada atributo, que é uma métrica que computa o quão relevante cada atributo é para a predição do fenômeno [6].

Por fim, para a implementação, o código de estratégia da equipe foi modificado de forma que sempre que algum robô se aproxima da bola, ele avalia a jogada de realizar o passe para cada um de seus aliados e de chutar ao gol e, com base nisso, decide sua jogada.

4. Resultados

Enquanto *Gradient Boosting* e *w-kNN* foram os melhores modelos de predição de chute com R^2 de teste em torno de 0,19, *Floresta Aleatória* e *AdaBoost* obtiveram 0,16 e *kNN* e *Regressão Linear* 0,13. Os modelos de predição de passe, no entanto obtiveram resultados próximos neste quesito, com todos os modelos obtendo R^2 de teste próximo de 0,21.

Calculando a importância dos atributos dos modelos gerados com base no conjunto de dados, cerca de 60% da avaliação de uma jogada de chute foi devida à distância da bola ao gol adversário, enquanto os restantes 40% foram devidos à distância do adversário mais próximo à bola e ao ângulo livre para chute. Na predição de passe, o ângulo livre de passe foi por muito o atributo mais importante; a distância do passe e a marcação sobre o passador e o receptor também foram relevantes.

Os modelos mais rápidos foram os de *Regressão Linear*, com t_{pred} próximo a 30ns. Os modelos de *Ensemble Learning* obtiveram t_{pred} próximo a 1 μ s;

enquanto os modelos (*w-kNN*) foram os mais lentos, com t_{pred} entre 3 μ s e 40 μ s na predição de chute e t_{pred} de passe entre 14 μ s e 90 μ s.

5. Conclusões

O melhor modelo de predição de chute obteve R^2 próximo a 0,19, isso significa que o modelo obtido é capaz de explicar 19% da variação da avaliação de uma jogada de chute a gol enquanto os demais 81% representam a variação não justificada pelo modelo. O mesmo é válido para os modelos de passe com R^2 próximo a 0,21.

Algumas imprecisões foram observadas durante o ajuste dos modelos, tais como ruído no conjunto de dados, atributos não considerados e principalmente pela incerteza do fenômeno, especialmente porque foram analisadas partidas de várias equipes que se comportavam de maneira diferente dentro do campo. Por conta destes fatores, os melhores modelos obtiveram R^2 de teste relativamente baixas. Entretanto, mesmos com os erros observados, com os modelos obtidos, já se tornou possível estimar a qualidade de uma jogada de passe ou chute em situações de jogo.

Dos métodos analisados, a *Regressão Linear* se provou ser um método rápido de implementação simples, porém inflexível na solução de problemas complexos. *kNN* se mostrou ser fácil de se implementar, porém bem lento, *w-kNN* aprimorou a qualidade de previsão do *kNN*, porém tornou o modelo final ainda mais lento. Os métodos de *Ensemble Learning*, como um todo, foram rápidos (perdendo apenas para os métodos lineares) e bastante ajustáveis, mas o ajuste dos seus vários hiperparâmetros foi cansativa.

De uma maneira geral, cada método teve pontos fracos e fortes, seja na praticidade de implementação ou no tempo e na qualidade das previsões do modelo final.

6. Referências

- [1] A. Burkov, "The hundred-page machine learning book". Andriy Burkov Canada, vol. 1. 2019.
- [2] A. C. Müller, S. Guido. "Introduction to Machine Learning with Python". O'Reilly Media, 2016.
- [3] J. Gou & L. Du & Y. Zhang & T. Xiong. "A New Distance-weighted k -nearest Neighbor Classifier". J. Inf. Comput. Sci.. 9, 2011.
- [4] S. Shalev-Shwartz, S. Ben-David. "Understanding Machine Learning: From Theory to Algorithms". Cambridge University Press. 2014.
- [5] N. J. D Nagekerke. "A Note on a General Definition of the Coefficient of Determination". Biometrika, [Oxford University Press, Biometrika Trust], v. 78, n. 3, p. 691–692, 1991.
- [6] G. LOUPPE. "Understanding Random Forests: From Theory to Practice". Tese (Doutorado). 2014.

Agradecimentos

À instituição FEI e à equipe RoboFEI pela realização das medidas e empréstimo de equipamentos e do laboratório para testes. À minha família pelo suporte, apoio e motivação.

¹ Aluno de IC do Centro Universitário FEI. Projeto com vigência de 08/2021 a 07/2022.