

UM ESTUDO SOBRE A APLICAÇÃO DE MÉTODOS DE APRENDIZADO DE MÁQUINAS PARA A PREDIÇÃO DE RESULTADOS EM FUTEBOL

Filipe Bosqueiro Bragil*, José Leonardo Takahashi, Leonardo Tomazeli Duarte

Resumo

O presente trabalho tem como objetivo investigar a aplicação de métodos de predição de aprendizado de máquinas no problema de previsão de resultados de jogos de futebol. Foi proposto um pré-processamento dos dados utilizando um Método Multicritério de Apoio a Decisão (AMCD), o PROMETHEE II. A abordagem foi validada com os dados dos jogadores da liga inglesa na temporada de 2016/17.

Palavras-chave:

Previsão de Resultados, Aprendizado de Máquinas, PROMETHEE III.

Introdução

Prever resultados é uma tarefa complexa devido à influência de uma grande quantidade de variáveis probabilísticas. Por outro lado, a formulação de técnicas de predição eficientes pode ter grande impacto dentro do universo esportivo, contribuindo, por exemplo, para tomadores de decisão estabelecerem estratégias com embasamento teórico. Diante deste panorama, o presente trabalho tem como objetivo realizar uma investigação acerca de métodos de predição de resultados.

Metodologia

Para traçar modelos de predição, primeiramente deve-se escolher as informações que serão utilizadas para se realizar a predição. Para o atual estudo, são considerados 25 critérios dos jogadores de futebol coletados da temporada da *Barclays Premier League* 2016/2017.

Com as informações selecionadas, além dos dados brutos reais, propomos a aplicação da média acumulada e a média variável por posição como *inputs*.

Após a definição das entradas ou *inputs*, é proposto uma etapa de pré-processamento, com subdivisões por rodadas, aplicando a ferramenta PROMETHEE II, uma ferramenta de apoio a decisão multicritério, para produzir um ordenamento medidor de qualidade dos jogadores.

Para finalizar o ordenamento, é gerado um valor de preferência a partir da análise multicritério, o qual, consideramos como um índice de qualidade do jogador.

Na etapa seguinte, as respectivas qualidades dos jogadores são alocadas em uma matriz, na qual cada índice representa as 26 posições possíveis de escalação, na respectiva posição ocupada pelo atleta na rodada. As posições não utilizadas naquele jogo são preenchidas com o valor 0.

Por fim, é preciso definir os métodos comparativos para previsão. Inicialmente, foi utilizada uma regressão múltipla. Em um segundo momento, construiu o modelo de previsão de rede neural baseado no algoritmo Perceptron com Validação Cruzada (Rede PVC).

Os métodos de previsão exigem as etapas denominadas treinamento, o qual pode ser supervisionado ou não, e validação. Na presente pesquisa, foi utilizado meia temporada (360 jogos) para essas etapas, sendo a proporção de 80% destes para treinamento e 20% de validação. O restante da temporada foi utilizado para prever os resultados do modelo proposto.

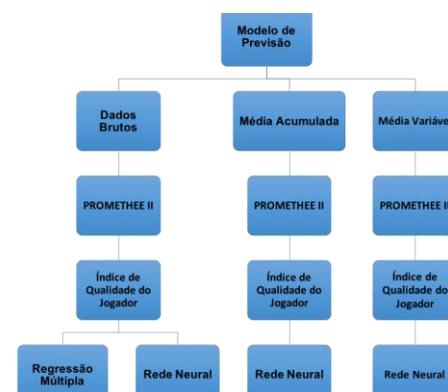


Figura 1. Resumo da Metodologia.

Resultados e Discussão

Com relação aos resultados obtidos, inicialmente, observamos o melhor desempenho durante o treinamento da Rede Neural comparado com a Regressão Múltipla.

Comparando o desempenho utilizando a Rede Neural durante o treinamento, percebe-se que tivemos um desempenho acima de 90% de acertos. Logo, as informações relevantes para previsão estão intrínsecas nos *inputs* da rede, o que é importante para o modelo preditor.

Com relação aos acertos da previsão, a rede neural com os *inputs* das médias (variável e acumulada) tiveram melhor desempenho comparado com a estratégia que utiliza unicamente os dados brutos.

Tabela 1. Resultados de Métodos e Entrada

Tipo de Dados de Entrada	Pré-Processamento	Método	Acertos Treinamento	Acertos Previsão
Dados Brutos	PROMETHEEII	Regressão Múltipla	0,46	0,41
Dados Brutos	PROMETHEEII	Rede PVC	0,93	0,47
Média Acumulada	PROMETHEEII	Rede PVC	0,96	0,66
Média Variável	PROMETHEEII	Rede PVC	0,911	0,761

Conclusões

Podemos concluir que o modelo de previsão com o melhor desempenho foi o com média variável. Assim como, utilizar um tratamento inicial das informações pode possibilitar um melhor desempenho no modelo preditor e, conseqüentemente, um dos caminhos para prever os resultados de futebol.