



UMA NOVA ABORDAGEM APLICADA AO CONCEITO MONEYBALL COM APOIO DA COMPOSIÇÃO PROBABILÍSTICA DE PREFERÊNCIAS

Luiz Octávio Gavião
luiz.gaviao67@gmail.com

Annibal Parracho Sant'Anna
annibal.parracho@gmail.com

Gilson Brito Alves Lima
Universidade Federal Fluminense
Rua Passo da Pátria, 156, Bloco D, São Domingos, Niterói-RJ, CEP 24.220-240
glima@id.uff.br

RESUMO

A recomposição de equipes com recursos limitados tornou-se um problema globalizado no mundo empresarial, nele inserido o contexto futebolístico. Nesse domínio particular do esporte, o modelo “Moneyball” vêm se tornando sinônimo da análise quantitativa à tomada de decisão para transferência de jogadores. Sua aplicação bem-sucedida em competitivo mercado gera, naturalmente, a proteção comercial dos algoritmos. Com a base nas características do “Moneyball”, esta pesquisa propôs uma abordagem probabilística de apoio à decisão multicritério. A pesquisa reuniu, de forma original, os algoritmos de classificação (CPP-TRI), de evolução de produtividade (Malmquist) e de regularidade de desempenho (Gini) da Composição Probabilística de Preferências (CPP). A proposta foi aplicada ao futebol, para avaliar a transferência de um jogador ao mercado europeu. Foram usados dados de desempenho e valor de mercado disponíveis em fontes abertas. Os resultados indicaram doze oportunidades de melhor investimento em jogadores do mesmo setor de atuação que o do contratado.

PALAVRAS CHAVE. CPP, Moneyball, Futebol, Contratação de Jogadores.

Tópicos: OA – Outras Aplicações; MP – Modelos Probabilísticos; SE – PO em Serviços

ABSTRACT

The recruitment of new athletes with limited resources became a globalized problem in sports. The experience of the Oakland Athletics’ manager turned the ‘Moneyball’ model as synonym of quantitative analysis in the transfer market of baseball players. However, the successful implementation in a competitive market forced the model to have its algorithms restricted to business users. Based on the ‘Moneyball’ characteristics, this research proposed a new probabilistic approach to support the decision-making process. The ‘Moneyball’ model is pruned to a probabilistic approach. Data collection depends on observers, who add imprecision to samples. In addition, a soccer match has a random component. The novelty here refers to the use of three algorithms explored by the Composition of Probabilistic Preferences: classification, the Malmquist index of productivity and the performance regularity. The new model analyzed a recent transfer of a soccer player to Europe. The results indicated twelve opportunities for better investment.

KEYWORDS. CPP. Moneyball. Soccer. Recruitment of players.

Paper topics - OA; MP; SE



1. Introdução

O mercado de transferências de jogadores de futebol envolve cifras bilionárias. Segundo a Federação Internacional de Futebol (FIFA), o Brasil foi o país que mais movimentou jogadores ao final de 2016. No período de um ano, a balança comercial do país indicou a “exportação” de 806 jogadores e “importação” de 678 atletas, envolvendo recursos aproximados em R\$ 1,88 bilhão [Mansur 2017]. Esse valor, equivalente à transferência de 1484 atletas, é inferior ao valor de mercado dos três times de futebol mais ricos do mundo: Real Madrid e Barcelona, da Espanha, e Manchester United, da Inglaterra [Ozarian 2016].

A realidade financeira da maioria dos clubes de futebol é diferente da dos que integram as Ligas mais valorizadas na Europa [Anderson & Sally 2013]. O campeonato brasileiro, por exemplo, ocupa a décima posição em valor de mercado dos jogadores, quando comparado aos torneios das Ligas Europeias [Barbosa 2014]. Os jogadores brasileiros que se destacam nos torneios pré-profissionais são precocemente negociados ao exterior, em face da dificuldade dos clubes em cobrir as ofertas de Ligas com maior capacidade de investimento [Cavalcanti & Capraro 2015]. Dessa forma, os clubes brasileiros precisam recompor seus elencos para as novas temporadas com limitados recursos.

O problema da recomposição de elencos com limitados recursos foi explorado por [Lewis 2004] em livro e, posteriormente, no filme “Moneyball”. Embora o contexto retratasse a realidade da Liga Americana de Baseball (MLB), a experiência do gestor do Oakland Athletics é extensiva a outras modalidades esportivas [Alamar & Mehrotra 2011]. A impossibilidade de cobrir as propostas milionárias de clubes tradicionais da MLB exigia uma nova abordagem na gestão da equipe. Era preciso identificar qualidades técnicas em jogadores mais baratos que, por algum motivo, não despertavam maior interesse de outros clubes. Desde então, diversos trabalhos acadêmicos tem explorado e adaptado o modelo “Moneyball” a outras modalidades esportivas: golfe e tênis [Baughman et al. 2015], hóquei [Mason 2006], basquete [Eldridge 2010], futebol americano [Whiting & Maynes 2016], futebol australiano [Stewart 2007] e ao próprio “soccer” [Tunaru & Viney 2010; Weimar & Wicker 2014].

A inclusão da análise quantitativa no apoio à decisão de compra e venda dos jogadores permitiu identificar atletas a partir do desempenho técnico nos fundamentos do esporte. Essa nova abordagem contrastava com os métodos tradicionais de escolha de jogadores. As reuniões com observadores de campo, ou “olheiros” no jargão esportivo, definiam as contratações com base em critérios subjetivos e parciais. Embora enfrentasse severa resistência do clube e da mídia especializada, a nova estratégia se mostrou bem-sucedida ao longo das temporadas no início dos anos 2000. O modelo de Billy Beane se tornou sinônimo da análise estatística em apoio à decisão dos gestores esportivos. Embora a concepção do modelo estivesse relacionada aos clubes de menor investimento, o Boston Red Sox, clube tradicional da MLB, também aplicou a abordagem com sucesso nas temporadas seguintes [Lewis 2004].

No caso do futebol brasileiro, o uso da análise quantitativa para a gestão do elenco de jogadores ainda é incipiente. Os clubes de maior investimento iniciam, aos poucos, a criação de departamentos para a análise estatística. Nas transmissões dos jogos de futebol, já é possível identificar sensores em coletes de tórax dos jogadores com maior frequência. A pesquisa científica tem crescido com a criação de cursos voltados à ciência do esporte em renomadas instituições de ensino no país e publicação de trabalhos acadêmicos [Fialho 2016; Ramos Filho & Ferreira 2016; Magalhães et al 2016].

O modelo “Moneyball” apresenta as características típicas de um problema de apoio à decisão multicritério (MCDA). Os métodos de MCDA oferecem soluções satisfatórias a problemas envolvendo alternativas avaliadas sob múltiplos critérios de desempenho. No caso sob análise, as alternativas são representadas por jogadores com potencial para a contratação, com avaliações de desempenho nos fundamentos técnicos do esporte e em seus valores de mercado.

A abordagem probabilística é especialmente útil ao modelo “Moneyball”. A coleta dos dados ainda requer a interferência de equipes de observadores e, dessa forma, é lícito admitir que as amostras reservam alguma imprecisão. Além disso, o caráter aleatório dos números em uma



partida de futebol confere maior aderência para a abordagem probabilística em relação a métodos determinísticos, conforme indicam as pesquisas de [Anderson & Sally 2013]. Porém, uma busca sistemática nas principais bases de pesquisa indica uma lacuna em relação ao uso de um modelo MCDA probabilístico, com base no conceito “Moneyball”. Assim, este artigo tem por objetivo apresentar uma proposta de adaptação do conceito “Moneyball” ao futebol, sob o ponto de vista MCDA probabilístico.

Esta pesquisa utilizou a Composição Probabilística de Preferências (CPP) para modelar o conceito “Moneyball”, doravante denominado “CPP-mb”, para abreviar o texto. O CPP foi desenvolvido por [Sant’Anna & Sant’Anna 2001], recebendo novas aplicações em [Sant’Anna 2015]. O CPP é um método MCDA probabilístico, cujas aplicações atendem às necessidades do problema. O CPP-mb reúne os algoritmos do CPP já utilizados em problemas de classificação, de análise da evolução de produtividade e de avaliação da regularidade das alternativas [Andrade & Sant’Anna 2013; Sant’Anna et al. 2012; Sant’Anna 2015]. A modelagem com os três algoritmos para a solução de um problema também é inédita na literatura do CPP.

Um caso recente de transferência de um jogador sul-americano ao mercado europeu foi analisado à luz do CPP-mb. Os dados são oriundos de fontes abertas, incluindo avaliações de desempenho técnico (i.e. WhoScored.com) e de valor de mercado (i.e. Transfermarkt.de). As fontes utilizadas nessa pesquisa tem sido exploradas em dissertações [Ellefsrød 2013; Klaiber 2016] e artigos científicos [Dendir 2016; Ruijg & van Ophem 2015]. Um conjunto pré-determinado de jogadores foi analisado. Por privacidade e termos de uso dos dados, a confidencialidade da amostra foi mantida. Em relação aos resultados, o CPP-mb indicou outras doze oportunidades de investimento em jogadores do mesmo setor de atuação e do mesmo mercado.

O artigo foi elaborado em cinco Seções. A Seção 2 efetua uma revisão bibliográfica, com ênfase ao CPP, com as aplicações de interesse indicadas. A Seção 3 apresenta a metodologia proposta nessa pesquisa. A Seção 4 aplica o CPP-mb para analisar a contratação de um jogador de futebol. Por fim, a Seção 5 traz as considerações finais.

2. Revisão Bibliográfica

2.1 O método CPP

O CPP é um método MCDA. Problemas que envolvem diferentes objetivos podem ser analisados por MCDA. De forma geral, esses objetivos estão representados por critérios de desempenho para a análise das diferentes alternativas à disposição do decisor. Por vezes, os critérios são conflitantes entre si, tornando complexa a escolha de uma alternativa em detrimento das demais [Pomerol & Barba-Romero 2012].

Em relação à abordagem probabilística, o método utiliza o conceito-chave da “aleatorização” das alternativas. Esse conceito deriva da imprecisão das medidas, da natureza aleatória do problema ou mesmo da incerteza acerca do contexto que envolve a tomada de decisão. Caso esses elementos não estivessem presentes, outros métodos determinísticos poderiam ser aplicados ao problema com soluções satisfatórias [Sant’Anna 2015]. Porém, o tratamento probabilístico de um problema sob tais condições confere maior realismo à modelagem.

Nesse primeiro passo do método CPP, as avaliações exatas das alternativas são interpretadas como medidas de posição (i.e. média, moda, mediana) de um conjunto de dados. Essa transformação permite o tratamento estatístico das avaliações por funções de probabilidade. A escolha da função probabilística que melhor se ajusta aos dados pode advir de informações *a priori*, de aplicações bem-sucedidas a problemas similares, do ajuste dos dados a distribuições conhecidas por técnicas estatísticas, denominado “goodness-of-fit” ou, na ausência dessas informações, arbitrada ao contexto. De maneira geral, o uso de distribuições normais e triangulares têm suprido as necessidades da modelagem simétrica e assimétrica dos dados,



respectivamente, com resultados satisfatórios, sendo ambas utilizadas nesta pesquisa [Sant'Anna 2015].

$$M_{ij} = \int_{D_{X_i}} \left[\prod F_{X_{-i}}(x_{-i}) \right] f_{X_i}(x_i) dx_i \quad (1)$$

$$m_{ij} = \int_{D_{X_i}} \left[\prod (1 - F_{X_{-i}}(x_{-i})) \right] f_{X_i}(x_i) dx_i \quad (2)$$

Após o processo de “aleatorização”, as probabilidades da *i*-ésima alternativa ser superior (M_{ij}) e inferior (m_{ij}) às demais, conjuntamente, são calculadas para cada *j*-ésimo critério. Nas equações (1) e (2), um vetor aleatório, “*X*”, simboliza as alternativas do problema. As funções relativas à *i*-ésima alternativa são indexadas por “*i*” e as demais por “*-i*”. As notações F_X , f_X e D_{X_i} são, respectivamente, a função distribuição cumulativa (cdf), a função densidade de probabilidade (pdf) e o domínio da variável aleatória “*X*” [Sant'Anna et al. 2012].

Para a ordenação das alternativas, as probabilidades “ M_{ij} ” e “ m_{ij} ” são compostas em escores de preferência global, para diferentes pontos de vista do decisor. As composições com base em “ M_{ij} ” são denominadas progressistas, por associar a decisão à maximização de ganhos. Nesse caso, os ganhos estão associados às probabilidades de maximizar as preferências. As composições com base em “ m_{ij} ” são denominadas conservadoras, por associar a decisão à minimização de perdas, representadas pelas probabilidades de minimizar as preferências. As composições que requerem a maximização “ M_{ij} ” e a minimização “ m_{ij} ” em todos os critérios são denominadas pessimistas, enquanto as otimistas são satisfeitas com a máxima ou mínima preferência em ao menos um critério. Dessa forma, é possível associar a tomada de decisão a quatro pontos de vista, referentes à interseção desses dois eixos progressista-conservador e otimista-pessimista.

Os cálculos das composições podem ainda considerar a independência ou máxima dependência das avaliações. Na hipótese mais simples de independência das avaliações, o decisor objetiva selecionar a alternativa com a melhor performance segundo o conjunto de critérios. Essa composição, denominada Progressista-Pessimista (PP), é obtida com o produto dos “ M_{ij} ” da *i*-ésima alternativa na totalidade dos critérios, com a seleção do maior escore global. A composição PP foi aplicada na Etapa 6 da pesquisa, conforme descrita na Seção 3, assumindo um comportamento normal para a “aleatorização” das avaliações. As fórmulas de cálculo das demais composições e para diferentes hipóteses de correlação das avaliações, não foram aplicadas nessa pesquisa e podem ser acessadas em [Sant'Anna 2015].

2.2 A classificação probabilística

O procedimento de classificação probabilística é uma variante do CPP. Sob o ponto de vista MCDA, a classificação das alternativas foi primeiramente desenvolvida por [Yu 1992] com o método ELECTRE TRI. O método CPP-TRI apresenta finalidade similar, porém com abordagem probabilística dos dados. Isto torna o CPP-TRI útil à solução de problemas MCDA em contextos de elevada incerteza e imprecisão dos dados [Sant'Anna et al. 2012].

A classificação das alternativas com o CPP-TRI requer a definição antecipada das classes. As classes são identificadas por perfis representativos. Para cada critério, esses perfis são referências para classificar uma alternativa em diferentes níveis de qualidade (i.e. alto, médio e baixo desempenho, dentre outras). Os perfis de classes podem ser estabelecidos com base em informações *a priori* do contexto, por escolha de especialistas no problema, por medidas de posição estatística, dentre outras possibilidades [Sant'Anna 2015]. No problema em questão foi usada a divisão em três classes, cada uma sendo representada por um único perfil. As três classes foram definidas pelos quantis de 65%, 50% e 35%, para representar jogadores com alto, médio e baixo desempenho nos fundamentos técnicos, respectivamente.

$$A_{ij}^+ = \prod_h P[X_j > Y_{ihj}] \quad (3)$$



$$A_{ij}^- = \prod_h P[X_j < Y_{ihj}] \quad (4)$$

Para efetuar a classificação, são calculadas as probabilidades de as alternativas “X” apresentarem valores acima “A+” e abaixo “A-” das pseudo-alternativas “Y”. Essas pseudo-alternativas são compostas com as avaliações correspondentes aos perfis das classes previamente estabelecidos. As Equações (3) e (4) descrevem os cálculos de “A+” e “A-”, para a i-ésima alternativa, j-ésimo critério e h-ésimo perfil. O menor valor absoluto da diferença entre essas probabilidades determina a classificação da alternativa. O cálculo das probabilidades de cada alternativa “X” receber preferência superior ou inferior às pseudo-alternativas “Y” é obtido com as Equações (1) e (2) [Sant’Anna 2015].

2.3 O Índice de Malmquist probabilístico

O Índice de Malmquist é normalmente utilizado para comparar a evolução de desempenho de unidades, serviços, recursos humanos, dentre outros, entre dois períodos considerados. O conceito original foi desenvolvido por [Malmquist 1953], com a finalidade de analisar a evolução do comportamento de consumidores. Posteriormente, o conceito foi explorado [Moorsteen 1961], tendo a evolução de produtividade aperfeiçoada por [Caves et al. 1982]. Depois, [Fare et al. 1994] adaptaram o conceito de [Malmquist 1953] ao modelo de Análise Envoltória de Dados (DEA). Recentemente, [Andrade & Sant’Anna 2013] adaptaram o índice ao cálculo probabilístico. Essa abordagem difere das demais por assumir a imprecisão das medidas, enquanto as demais versões exploram os dados de forma determinística.

O conceito-chave do índice de Malmquist está relacionado à avaliação dinâmica de alternativas segundo os mesmos critérios, para diferentes momentos. Dessa forma, é possível identificar a evolução de uma alternativa entre dois momentos “t” e “t+1”. Os cálculos do índice envolvem a relação entre as distâncias de uma alternativa para fronteiras de produção do conjunto de alternativas nesses dois momentos. Em DEA, essas distâncias equivalem às eficiências da alternativa avaliada sob os diferentes conjuntos de dados em cada momento.

Na abordagem probabilística do CPP, essas distâncias são probabilidades de as alternativas apresentarem a maior e menor eficiências. Assim, as alternativas são comparadas em relação às demais e não apenas com as fronteiras. Além disso, o índice probabilístico combina escores relativos aos extremos progressista e conservador, utilizados no CPP. Isto confere maior estabilidade ao cálculo do índice probabilístico [Sant’Anna 2015]. Além disso, o CPP tampouco apresenta as limitações de dimensionalidade entre a quantidade de alternativas avaliadas e de variáveis do problema. Por fim, a abordagem probabilística é não-linear, valorizando os desempenhos mais próximos às fronteiras, o que pode tornar a modelagem mais aderente a problemas reais.

O índice de Malmquist probabilístico pode ser calculado a partir das Equações (5) a (11). As probabilidades de uma alternativa apresentar a máxima ou mínima preferência em relação às demais foram anteriormente descritas nas Equações (1) e (2). Os fatores progressista (MP) e conservador (MC) do índice são calculados com base nessas probabilidades de maximizar e minimizar as preferências, respectivamente, em relação aos momentos “t” e “t+1”. Nas equações, a notação “a” representa a a-ésima alternativa, “j” o j-ésimo critério e “ π ” o produto dos termos. O fator conservador $MC(a,u,v)$, para a a-ésima alternativa, indica a sua probabilidade de não receber a mínima preferência por nenhum critério, com as suas avaliações no instante “u” e as das demais alternativas no instante “v”. O fator progressista $MP(a,u,v)$, para a a-ésima alternativa, indica a probabilidade de a a-ésima alternativa não receber a máxima preferência por nenhum critério, considerando os momentos “u” e “v” já descritos.

$$MP_{pré} = MP(a,t+1,t)/MP(a,t,t) = \pi_j(1-\pi_i P[X_{ajt+1} \geq X_{ijt}])/\pi_j(1-\pi_i P[X_{ajt} \geq X_{ijt}]) \quad (5)$$

$$MP_{pós} = MP(a,t+1,t+1)/MP(a,t,t+1) = \pi_j(1-\pi_i P[X_{ajt+1} \geq X_{ijt+1}])/\pi_j(1-\pi_i P[X_{ajt} \geq X_{ijt+1}]) \quad (6)$$

$$MP = (MP_{pré} * MP_{pós})^{1/2} \quad (7)$$



$$MC_{pré} = MC(a,t+1,t)/MC(a,t,t) = \pi_j(1-\pi_i P[X_{ajt+1} \leq X_{ijt}]) / \pi_j(1-\pi_i P[X_{ajt} \leq X_{ijt}]) \quad (8)$$

$$MC_{pós} = MC(a,t+1,t+1)/MC(a,t,t+1) = \pi_j(1-\pi_i P[X_{ajt+1} \leq X_{ijt+1}]) / \pi_j(1-\pi_i P[X_{ajt} \leq X_{ijt+1}]) \quad (9)$$

$$MC = (MC_{pré} * MC_{pós})^{1/2} \quad (10)$$

$$M = (MC / MP)^{1/2} \quad (11)$$

Assumindo a independência das avaliações dos critérios, o $MP_{pré}$ e $MP_{pós}$ da a -ésima alternativa serão dados pelo produto das probabilidades não ser a melhor, conforme as Equações (5) e (6). O uso do produto contribui para reduzir a influência de valores extremos. O MP é obtido a partir da média geométrica entre o $MP_{pré}$ e $MP_{pós}$, conforme a Equação (7). Os cálculos do índice conservador MC seguem a mesma lógica do índice MP, sendo descritos nas Equações (8) a (10). Finalmente, o índice de Malmquist probabilístico (M) pode ser obtido com a média geométrica dos índices MP e MC, conforme a Equação (11).

2.4 O Índice de Gini probabilístico

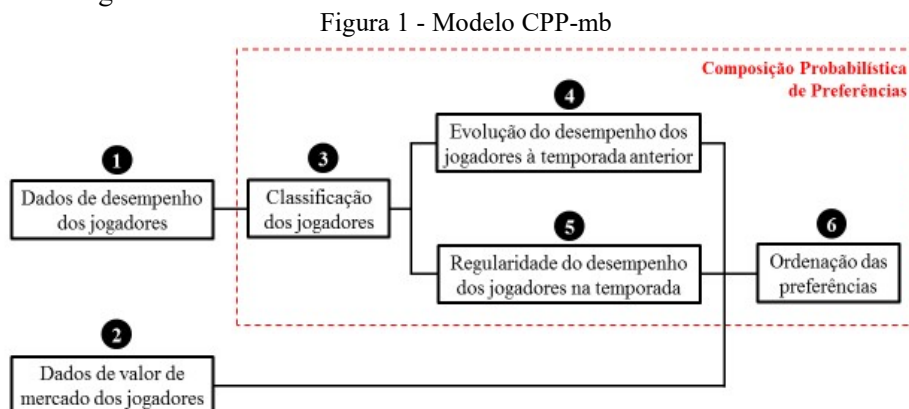
O índice de Gini foi desenvolvido por [Gini 1921], para identificar a desigualdade em um conjunto de valores. Esse índice tem sido normalmente usado para avaliar a desigualdade de renda de países. O índice de Gini pode variar entre “0” e “1”, em que o maior valor representa a máxima desigualdade entre os valores. Dessa forma, o índice é utilizado como medida aproximada das condições socioeconômicas de uma população [Hoffmann 2016]. Por exemplo, se todo componente de uma população perceber o mesmo salário, o índice de Gini para essa população é nulo.

$$G = \frac{1}{n} \left(n+1 - 2 \frac{\sum_{i=1}^n (n+1-i)y_i}{\sum_{i=1}^n y_i} \right) \quad (12)$$

O índice de Gini pode ser calculado para uma amostra de “n” dados, a partir da Equação (12) [Dixon et al. 1987]. A notação “ y_i ” representa os valores da amostra, sendo dispostos de forma crescente ($y_i \leq y_{i+1}$), em que $i \in \{1; n\}$. Na abordagem probabilística, os valores utilizados no cálculo do índice de Gini são as probabilidades “ M_{ij} ” e “ m_{ij} ”. Dessa forma, é possível ordenar as alternativas por regularidade do desempenho probabilístico nos diferentes critérios, a partir do método CPP.

3. Metodologia

O modelo CPP-mb foi conceitualmente fundamentado na abordagem utilizada por Billy Beane para a contratação de jogadores, considerando a análise quantitativa do desempenho e do valor de mercado dos atletas. O modelo proposto é desenvolvido em seis etapas, conforme apresentado na Figura 1.





As Etapas 1 e 2 envolvem, respectivamente, a coleta de dados de desempenho e de valor de mercado dos jogadores. Inicialmente devem ser relacionados os potenciais jogadores que preencham os requisitos estabelecidos para a tomada de decisão. É possível que somente jogadores de uma determinada posição sejam necessários para recompor o time, ou que atendam a determinada faixa-etária, por exemplo. Além disso, os critérios de desempenho julgados relevantes ao apoio à decisão devem ser submetidos aos responsáveis pela tomada de decisão. Esses aspectos qualitativos funcionam como filtros para a modelagem da matriz de decisão, composta por jogadores, critérios de desempenho e avaliações correspondentes. Essa matriz deve anteceder a aplicação do “CPP-mb”, como em qualquer problema MCDA [Pomerol & Barba-Romero 2012]. Os dados coletados na Etapa 2 são utilizados posteriormente no CPP-mb. Esses dados são utilizados para compor outra matriz de decisão, com os resultados das Etapas 4 e 5.

Na Etapa 3 efetua-se a classificação dos jogadores, preferencialmente em classes de alto, médio e baixo rendimento. Dependendo da quantidade de atletas para a avaliação, é possível adaptar o procedimento de cálculo, reduzindo a quantidade de classes. No caso de duas classes, por exemplo, seria possível considerar a segmentação dos jogadores em uma classe superior e outra inferior. A classificação antecede as demais etapas do CPP-mb, de forma a garantir que os resultados das etapas posteriores envolvam atletas com desempenhos similares. Isto evita o risco da escolha de um atleta de baixo rendimento global.

Na Etapa 4 efetua-se o cálculo da evolução do desempenho dos atletas por classe, entre dois períodos considerados. Os períodos podem ser escolhidos conforme o contexto de decisão. Assim, é possível coletar dados ao final das temporadas, dos turnos ou mesmo entre segmentos específicos da mesma temporada. Essa flexibilidade pode ser útil para antecipar a análise dos jogadores em relação às datas das “janelas” de transferência dos atletas, normalmente a cada semestre.

Na Etapa 5, os dados de desempenho são utilizados para calcular a regularidade do atleta. A opção por essa característica do jogador decorre da flexibilidade proporcionada aos treinadores. Um elenco de jogadores que dominam diferentes fundamentos em alto desempenho permite que o treinador explore maior variedade de esquemas táticos na equipe. Por fim, na Etapa 6 os jogadores são ordenados. A maior pontuação indica o atleta de melhor custo-benefício entre o valor de mercado e a análise de desempenho. Os cálculos das Etapas 3 a 6 são probabilísticos, a partir da modelagem do CPP.

Além da abordagem probabilística, o CPP-mb apresenta o uso de novas dimensões de desempenho técnico dos jogadores. Nessa pesquisa, o resultado é calculado a partir de quatro aspectos: primeiro, a classificação dos jogadores em subconjuntos de desempenho; segundo, a evolução de produtividade do jogador, entre as temporadas, turnos ou demais períodos considerados para análise; terceiro, a regularidade do jogador nos diferentes critérios de avaliação técnica; por fim, o valor de mercado do jogador. À exceção dos dados de valor de mercado, os três fatores técnicos são menos perceptíveis aos analistas qualitativos. Esses fatores requerem um “olhar” mais atento do avaliador e cálculos complexos para classificar, destacar a evolução em diferentes períodos e observar a regularidade dos desempenhos. Em resumo, o jogador que representa a melhor opção de investimento, à luz do CPP-mb, pertence ao subconjunto de alto desempenho técnico e apresenta o melhor custo-benefício em relação à evolução entre temporadas, à regularidade de desempenho e ao valor de mercado.

4. Aplicação

O CPP-mb foi utilizado para analisar uma contratação no mercado do futebol. O futebol sul-americano revela promissores atletas com elevada frequência e se tornou um celeiro de jogadores para os clubes com maior capacidade de investimento. O caso real aqui estudado se refere à transferência de um jogador sul-americano para o mercado europeu. Aplicou-se o CPP-mb para analisar essa contratação, em relação a um conjunto de atletas do mesmo setor de atuação do contratado. Foi possível constatar que outros doze jogadores apresentaram resultados mais satisfatórios que o do jogador contratado.



4.1 Coleta de dados e modelagem

A amostra de jogadores foi selecionada com base na posição do jogador contratado. Um conjunto de jogadores que atuaram no mesmo setor de campo, nas competições do país sul-americano, compôs a amostra. Tendo em vista a necessidade de avaliar o desempenho dos jogadores em dois momentos distintos, conforme a Etapa 4 do modelo proposto, estabeleceu-se a condição de coletar somente os jogadores que atuaram nas mesmas competições nas temporadas de 2015 e 2016. Ao final, a amostra conteve 32 jogadores para a análise. O jogador 12 foi o atleta contratado e consiste, dessa forma, na referência para comparar os resultados do CPP-mb aos demais jogadores.

Os dados de desempenho e valor de mercado dos jogadores foram coletados a partir de fontes abertas, nos sites WhoScored.com e Transfermarkt.com, respectivamente. A WhoScored.com se dedica à análise e desenvolvimento de software aplicado ao futebol. Os dados de desempenho dos jogadores são coletados em diferentes centros esportivos, com a cobertura de aproximadamente quinze mil equipes de futebol. A Transfermarkt.com disponibiliza dados de valor de mercado dos jogadores. Embora as estimativas envolvam informações comerciais e contratuais de acesso restrito, os resultados são considerados confiáveis no mercado do futebol [Bryson et al. 2013]. Entretanto, a abordagem probabilística do modelo aqui proposto permite lidar com a imprecisão das medidas e incerteza acerca da padronização na coleta e manejo dos dados. Por confidencialidade, a base de dados foi omitida.

Os dados da amostra foram coletados a partir de vinte critérios de desempenho individual. As avaliações foram padronizadas à quantidade de minutos de participação de cada jogador na temporada. As seguintes variáveis foram consideradas no estudo: gols marcados, gols contra, assistências a gols, ações contra oportunidades de gol do adversário, chutes a gol, passes terminais para chutes a gol, percentual de passes certos, passes longos, bloqueio de chutes do adversário, cartões disciplinares, cruzamentos, desarmes efetuados, desarmes sofridos, interceptações de passes, faltas cometidas, faltas sofridas, impedimentos de ataque, impedimentos impostos ao adversário, dribles efetuados e dribles sofridos.

Houve jogadores que apresentaram desempenho nulo em alguns critérios. Esses dados nulos foram transformados em unidades de milésimo, de forma que a diferença de tempo de participação nas partidas diferenciasse o desempenho dos jogadores. Por exemplo, dois jogadores “A” e “B”, com desempenhos nulos em um dado critério e tempos de participação de 200 e 300 minutos, teriam seus dados transformados em $5E-6$ e $3,3E-6$, respectivamente. O resultado indica o melhor desempenho do jogador “A”, pois o desempenho nulo foi registrado em um tempo menor de participação nas partidas.

Os vinte critérios receberam pesos iguais. Atribuir relevância a alguns critérios em detrimento de outros pode mascarar a realidade do jogo e inverter resultados. Segundo [Anderson & Sally 2013], critérios ofensivos tendem a ser supervalorizados, porém as ações defensivas que evitam gols desempenham papel decisivo em jogos. Um desarme efetuado em um momento chave da partida pode ter evitado uma derrota, por exemplo. A imprevisibilidade no futebol é elevada, relativamente às demais modalidades esportivas e cada critério pode influenciar ou desencadear eventos de maneira aleatória em uma partida. Esta é uma razão para preferirmos a abordagem de composição dos critérios pela maximização conjunta das preferências, que prescinde de pesos e atribui igual importância a todos critérios.

4.2 Análise dos resultados

Após a composição da matriz de decisão, com jogadores, critérios e avaliações, foi efetuada a classificação dos jogadores. Posteriormente, foram efetuados os cálculos dos Índices de Malmquist e de Gini, para cada classe. Em seguida, o ponto de vista PP do CPP foi aplicado a esses dois índices, juntamente com o valor de mercado dos jogadores, para a ordenação final. Esse ponto de vista conduz a garantir a escolha de alternativa com bom desempenho em relação a



esses três critérios simultaneamente. Os cálculos correspondentes foram descritos na Seção 2.3, sendo efetuados com auxílio do software R [R-Core-Team 2016]. A Tabela 1 apresenta os resultados da Etapa 3 com aleatorização por distribuições triangulares para as classes de alto e médio desempenho, que reuniram nove e 17 jogadores, respectivamente, sendo omitidos os resultados da classe de baixo desempenho. Os números da coluna “Jogador” substituem seus nomes.

Tabela 1 - Resultados das classes de alto e médio desempenho

| Jogador | Classe | Índice de Malmquist | R1 | Índice de Gini | R2 | Valor de Mercado | R3 | CPP (PP) | Resultado |
|---------|--------|---------------------|----|----------------|----|------------------|----|----------|-----------|
| 20 | 3 | 2,6218 | 1 | 0,3677 | 3 | 1,5 | 1 | 1,30E-02 | 1 |
| 30 | 3 | 2,5341 | 2 | 0,3785 | 4 | 2 | 4 | 5,01E-03 | 2 |
| 8 | 3 | 1,5368 | 4 | 0,3410 | 1 | 3 | 7 | 4,13E-04 | 3 |
| 27 | 3 | 1,2923 | 6 | 0,3575 | 2 | 2,5 | 5 | 2,76E-04 | 4 |
| 32 | 3 | 1,0525 | 7 | 0,3895 | 6 | 1,5 | 1 | 1,77E-04 | 5 |
| 6 | 3 | 0,8304 | 8 | 0,3860 | 5 | 1,5 | 1 | 9,34E-05 | 6 |
| 7 | 3 | 1,3579 | 5 | 0,3978 | 7 | 3,5 | 8 | 2,87E-05 | 7 |
| 21 | 3 | 0,6263 | 9 | 0,5096 | 9 | 2,5 | 5 | 1,41E-07 | 8 |
| 15 | 3 | 2,2330 | 3 | 0,5060 | 8 | 6 | 9 | 2,88E-08 | 9 |
| Jogador | Classe | Índice de Malmquist | | Índice de Gini | | Valor de Mercado | | CPP (PP) | Resultado |
| 24 | 2 | 0,9878 | 1 | 0,1884 | 1 | 1,5 | 8 | 4,33E-03 | 1 |
| 5 | 2 | 0,5686 | 12 | 0,3041 | 4 | 0,75 | 2 | 6,79E-05 | 2 |
| 22 | 2 | 0,8567 | 3 | 0,3685 | 8 | 1,5 | 8 | 6,73E-05 | 3 |
| 12 | 2 | 0,7330 | 5 | 0,1961 | 2 | 2,5 | 13 | 4,31E-05 | 4 |
| 31 | 2 | 0,4938 | 14 | 0,3103 | 5 | 0,5 | 1 | 3,53E-05 | 5 |
| 11 | 2 | 0,7420 | 4 | 0,4138 | 11 | 0,75 | 2 | 3,38E-05 | 6 |
| 10 | 2 | 0,6707 | 7 | 0,3772 | 10 | 1 | 4 | 2,78E-05 | 7 |
| 18 | 2 | 0,9722 | 2 | 0,4793 | 15 | 1 | 4 | 1,91E-05 | 8 |
| 2 | 2 | 0,6547 | 8 | 0,2975 | 3 | 2 | 11 | 1,50E-05 | 9 |
| 29 | 2 | 0,6126 | 10 | 0,3114 | 6 | 2 | 11 | 7,27E-06 | 10 |
| 25 | 2 | 0,4939 | 13 | 0,3699 | 9 | 1 | 4 | 4,20E-06 | 11 |
| 4 | 2 | 0,6747 | 6 | 0,4251 | 12 | 1,75 | 10 | 1,76E-06 | 12 |
| 16 | 2 | 0,4746 | 15 | 0,4579 | 14 | 1 | 4 | 3,02E-07 | 13 |
| 23 | 2 | 0,6404 | 9 | 0,4340 | 13 | 2,5 | 13 | 1,34E-07 | 14 |
| 28 | 2 | 0,3592 | 16 | 0,3656 | 7 | 2,75 | 15 | 9,45E-09 | 15 |
| 26 | 2 | 0,5764 | 11 | 0,4860 | 16 | 3,5 | 16 | 4,72E-10 | 16 |
| 17 | 2 | 0,1810 | 17 | 0,5214 | 17 | 3,5 | 16 | 3,44E-13 | 17 |

Fonte: Autores.

A Tabela 1 está ordenada por preferência, expressa na coluna “Resultado”, para as classes de alto e médio desempenho. O Jogador 20 representa a melhor opção de investimento. As colunas “R1” e “R3” indicam que o Índice de Malmquist probabilístico e o valor de mercado do Jogador 20 foram o maior e menor, respectivamente, enquanto obteve a terceira posição no Índice de Gini. A evolução técnica desse jogador, entre as temporadas de 2015 e 2016, a regularidade de desempenho nos critérios estabelecidos e o valor de mercado foram computados no CPP, sob o ponto de vista PP.

O Jogador 12, que fora contratado pelo clube europeu, apresentou a quarta colocação final, na classe de médio desempenho. Cabe destacar que esse jogador obteve a segunda melhor



regularidade em sua classe, demonstrando um equilíbrio no desempenho quanto aos vinte critérios em 2016. Seus resultados indicam que foi o quinto em evolução de desempenho entre as temporadas 2015-2016, na classe intermediária.

Com base no conceito “Moneyball”, o Jogador 12 não representou a melhor opção de investimento. Sob o ponto de vista quantitativo, outras opções apresentaram melhores resultados. De fato, um modelo matemático não pode garantir que o Jogador 20 seja melhor sucedido que o Jogador 12, na próxima temporada. Afinal, uma série de aspectos interferem na adaptação do atleta a outro país, no estilo de jogo da equipe, na interação com os demais companheiros, dentre outros, que contribuirão para o sucesso ou fracasso do investimento.

Entretanto, a abordagem probabilística foi capaz de indicar o atleta de maior potencial e de melhor custo-benefício ao investidor. Mesmo que outros aspectos qualitativos indicassem a contratação do Jogador 12, em detrimento dos demais atletas analisados, os gestores teriam reduzida a subjetividade da escolha. Dentre os doze jogadores com melhor potencial que o contratado, nove foram classificados com melhor desempenho técnico nos fundamentos escolhidos para a análise e três foram classificados com desempenho intermediário, na mesma classe a que pertence o jogador contratado.

5. Considerações Finais

O modelo “Moneyball” foi utilizado para analisar o mercado de transferência de jogadores de futebol. O modelo foi aplicado com sucesso para a recomposição da equipe do Oakland Athletics, da MLB. Na ocasião, a equipe apresentava sérias limitações financeiras para repor a perda dos seus melhores atletas para os clubes mais tradicionais do baseball americano. O modelo “Moneyball” se tornou mundialmente conhecido a partir do livro de [Lewis 2004] e de sua adaptação ao cinema. Em linhas gerais, o modelo priorizava aspectos quantitativos do desempenho dos jogadores e de seus valores de mercado, em contraste com as opiniões subjetivas dos “olheiros” do time. Desde então, o modelo “Moneyball” tem recebido adaptações em diferentes esportes.

Nessa pesquisa, utilizou-se uma abordagem MCDA probabilística para modelar o “Moneyball”. A natureza e contexto do problema potencializam a eficiência desse tipo de abordagem. As pesquisas de [Anderson & Sally 2013] revelaram diversos estudos estatísticos que descrevem o caráter aleatório dos números em uma partida de futebol. Além disso, a coleta dos dados de desempenho técnico e valor de mercado ainda requer a interferência humana e, dessa forma, é lícito admitir que as amostras reservam alguma imprecisão. Os resultados obtidos com a abordagem probabilística foram satisfatórios. O CPP-mb contraindicou a contratação do jogador 12. Outros jogadores apresentaram melhores resultados, segundo os critérios avaliados.

O modelo proposto apresenta limitações, inerentes aos métodos MCDA e específicas ao CPP-mb. Os métodos MCDA, em geral, são sensíveis à escolha dos critérios. Isto significa que uma escolha inadequada dos fundamentos técnicos levará a resultados não confiáveis. Outro aspecto importante a ressaltar se refere à finalidade de apoio à decisão dos métodos MCDA. Isto reflete o papel de suporte do método MCDA e não à decisão propriamente dita. Segundo [Anderson & Sally 2013], o equilíbrio na tomada de decisão é essencial e os resultados da aplicação do MCDA representam apenas um fator da escolha. Em relação ao CPP-mb, cabe destacar que a revisão da literatura não indica qualquer aplicação ao “Moneyball” com similaridade à natureza probabilística e MCDA do modelo proposto. Isto compromete a comparação do CPP-mb com modelos semelhantes, de forma a comprovar a sua utilidade.

A pesquisa pode ser estendida em estudos futuros. Dois trabalhos são visualizados. Primeiramente, o modelo CPP-mb é aplicável a outros esportes, cabendo somente o levantamento dos critérios e avaliações correspondentes à modalidade esportiva. Uma segunda variante de pesquisa está relacionada à convocação de jogadores para as seleções nacionais. Nesse contexto, a Etapa 2 deve ser desconsiderada, cabendo somente prosseguir no modelo com as variáveis e cálculos referentes ao desempenho técnico. Essa opção de pesquisa é interessante, pois na convocação de uma seleção está subentendido que o momento técnico do jogador é o fator mais



importante. Entretanto, uma série de jogadores convocados são questionados, pois aparentemente não apresentam o melhor desempenho em sua posição. Assim, o CPP-mb pode ser utilizado para comparar o jogador convocado com outros de sua mesma posição.

Referências

- Alamar, B., & Mehrotra, V. (2011). Beyond “Moneyball”: The rapidly evolving world of sports analytics, Part I. *Analytics Magazine*.
- Anderson, C., & Sally, D. (2013). Os números do jogo. Por que tudo o que você sabe sobre futebol está errado. *Editora Paralela. São Paulo*.
- Andrade, G. N. de, & Sant’Anna, A. P. (2013). Composição probabilística e índice de Malmquist para avaliação de eficiência em distribuidoras de energia elétrica. In *Simpósio Brasileiro de Pesquisa Operacional XLV SBPO* (Vol. 2013, pp. 937–948). Natal.
- Barbosa, D. (2014, April). Os 15 campeonatos de futebol mais valiosos do mundo. *Exame*. Retrieved from <http://exame.abril.com.br/negocios/os-25-campeonatos-de-futebol-mais-valiosos-do-mundo/>
- Baughman, A. K., Bogdany, R. J., McAvoy, C., Locke, R., O’Connell, B., & Upton, C. (2015). Predictive cloud computing with big data: Professional golf and tennis forecasting [application notes]. *IEEE Computational Intelligence Magazine*, 10(3), 62–76.
- Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent: footedness and player remuneration in European soccer. *Journal of Sports Economics*, 14(6), 606–628.
- Cavalcanti, E. A., & Capraro, A. M. (2015). Transferências internacionais no futebol: um estudo de caso comparativo entre os maiores clubes europeus e brasileiros. *RBFF-Revista Brasileira de Futsal E Futebol*, 7(23), 3–15.
- Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*, 1393–1414.
- Dendir, S. (2016). When do soccer players peak? A note. *Journal of Sports Analytics*, 2(2), 89–105.
- Dixon, P. M., Weiner, J., Mitchell-Olds, T., & Woodley, R. (1987). Bootstrapping the Gini coefficient of inequality. *Ecology*, 68(5), 1548–1551.
- Eldridge, R. (2010). *Measuring Efficiency in the National Basketball Association: A “Moneyball” Approach [Master Dissertation]*. NYU - Stern School of Business.
- Ellefsrød, M. B. (2013). *The Betting Machine: Using in-depth match statistics to compute future probabilities of football match outcomes using the Gibbs sampler [Master Thesis]*. Norwegian University of Science and Technology.
- Fare, R., Grosskopf, S., & Lovell, C. A. K. (1994). *Production frontiers*. Cambridge University Press.
- Fialho, G. (2016). Cursos de graduação e mestrado em Gestão do Esporte são apresentados na Universidade Federal Fluminense. Retrieved from <http://www.esporte.gov.br/index.php/ultimas-noticias/209-ultimas-noticias/55646-cursos-de-graduacao-e-mestrado-em-gestao-do-esporte-sao-apresentados-na-universidade-federal-fluminense>
- Ramos Filho, L. A. de O. R., & Ferreira, M. P. (2016). Gestão na formação de atletas: proposta de avaliação técnica para o futebol. *Revista Conbrad*, 1(3), 238–253.
- Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal*, 31(121), 124–126.
- Hoffmann, R. (2016). Desigualdade da distribuição da renda no Brasil: a contribuição de aposentadorias e pensões e de outras parcelas do rendimento domiciliar per capita. *Economia E Sociedade*, 18(1), 213–231.
- Klaiber, J. D. (2016). *Soccer Player Performance Rating Systems for the German Bundesliga [Master Thesis]*. Ghent University.



- Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.
- Magalhães, L. B., Castroneves, T., Chaves, M. C. C., Gomes, C. F. S., & Pereira, E. R. (2016). Estudo de apoio à decisão: a escolha do “Camisa 10” ideal baseado no método MacBeth. *Revista Brasileira de Futsal E Futebol*, 8(29), 113–128.
- Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadística Y de Investigación Operativa*, 4(2), 209–242.
- Mansur, C. E. (2017, February 3). Brasil é o recordista em exportação e importação de jogadores. *O Globo*. Rio de Janeiro. Retrieved from <http://oglobo.globo.com/esportes/brasil-o-recordista-em-exportacao-importacao-de-jogadores-20866699>
- Mason, D. S. (2006). Moneyball as a supervening necessity for the adoption of player tracking technology in professional hockey. *International Journal of Sports Marketing and Sponsorship*, 8(1), 41–55.
- Moorsteen, R. H. (1961). On measuring productive potential and relative efficiency. *The Quarterly Journal of Economics*, 75(3), 451–467.
- Ozanian, M. (2016, November 5). The Business Of Soccer. *Forbes*. Retrieved from <http://www.forbes.com/soccer-valuations/>
- Pomerol, J.-C., & Barba-Romero, S. (2012). *Multicriterion decision in management: principles and practice* (Vol. 25). New York: Springer.
- R-Core-Team. (2016). R: A language and environment for statistical computing. [Http://www. R-Project.Org](Http://www.R-Project.Org). Vienna, Austria.
- Ruijg, J., & van Ophem, H. (2015). Determinants of football transfers. *Applied Economics Letters*, 22(1), 12–19.
- Sant’Anna, A. P. (2015). *Probabilistic Composition of Preferences, Theory and Applications*. New York: Springer.
- Sant’Anna, A. P., Costa, H. G., & Pereira, V. (2012). CPP-TRI: um método de classificação ordenada baseado em composição probabilística. *Relatórios de Pesquisa Em Engenharia de Produção (UFF)*, 12(8), 104–117.
- Sant’Anna, A. P., Gomes, L. F. A. M., Costa, F. F. da, Rangel, L. A. D., Faria, M. J. da S., Ferreira, R. G., ... Senna, V. de. (2012). Análise multicritério baseada em probabilidades de preferência. In V. F. de Oliveira, V. Cavenaghi, & F. S. Másculo (Eds.), *Tópicos emergentes e desafios metodológicos em Engenharia de Produção: casos, experiências e proposições - Volume V* (p. 258). Rio de Janeiro: ABEPRO.
- Sant’Anna, A. P., & Sant’Anna, L. A. F. P. (2001). Randomization as a stage in criteria combining. In *International Conference on Industrial Engineering and Operations Management - VII ICIEOM* (pp. 248–256). Salvador.
- Stewart, M., Mitchell, H., & Stavros, C. (2007). Moneyball applied: Econometrics and the identification and recruitment of elite Australian footballers. *International Journal of Sport Finance*, 2(4), 231–248.
- Tunaru, R., & Viney, H. P. (2010). Valuations of soccer players from statistical performance data. *Journal of Quantitative Analysis in Sports*, 6(2), 410–1559.
- Weimar, D., & Wicker, P. (2014). Moneyball Revisited Effort and Team Performance in Professional Soccer. *Journal of Sports Economics*, 1527002514561789.
- Whiting, S. W., & Maynes, T. D. (2016). Selecting team players: Considering the impact of contextual performance and workplace deviance on selection decisions in the National Football League. *Journal of Applied Psychology*, 101(4), 484.
- Yu, W. (1992). *ELECTRE TRI (aspects méthodologiques et manuel d’utilisation)*. Document- Université de Paris-Dauphine, LAMSADE.