# Inteligência computacional aplicada à previsão de vencedores em partidas de tênis

Mateus de Araujo Fernandes<sup>1</sup>

Resumo: a previsão de vencedores em partidas de tênis pode representar várias utilidades práticas, pois os resultados de uma rodada em um torneio determinam quais jogos ocorrerão na rodada seguinte, o que é valioso para a organização dos torneios e a mídia, auxiliando na alocação de jogos em quadras e horários mais propícios, permitindo previsões de público e audiência e até mesmo embasando ações de merchandising. Neste trabalho, são estudados alguns dos principais fatores de influência na previsibilidade de partidas e, a partir dessa análise, são propostas duas diferentes abordagens para cálculos de confiabilidade na vitória de cada um dos competidores antes do início de uma partida. A primeira baseia-se em um sistema de inferência Fuzzy, explorando sua capacidade de reprodução de conhecimento de um especialista diante de uma mescla de informações. A segunda emprega uma rede neural, com sua característica de extração de atributos por meio de exemplos. Ambos os preditores têm como entradas dados de desempenhos prévios dos jogadores, que, nesse caso, tentam captar suas performances de curto, médio e longo prazo, além de sua afinidade com os diferentes tipos de pisos. Os resultados obtidos são encorajadores, mostrando ganhos significativos em relação à simples comparação baseada no ranking de entradas da Associação de Tenistas Profissionais.

Palavras-chave: Lógica Fuzzy. Predição. Redes neurais.

**Abstract:** The forecast of winners in tennis matches may represent several practical utilities, as the results of a round in a tournament determine which matches will occur in the next round. That is valuable to the tournaments' organizers and the media, assisting in the allocation of the matches in courts that are more appropriate and in the schedules, enabling public and audience forecasts and even basing merchandising actions. Here, we studied some of the main factors influencing matches predictability and, from this analysis, we propose two different approaches to calculations of chances of victory of each of the competitors before the start of a match. The first is based on a Fuzzy Inference System, exploring its ability to reproduce knowledge of an expert among mixed informations. The second employs a Neural Network, with its capability of features extraction from examples. Both predictors have as inputs data from previous performances of the players, that in this case try to capture their short, medium and long-term performances, as well as their affinity for the different types of surfaces. The results are encouraging, showing significant gains comparing to the simple use of the Association of Tennis Professionals ranking.

Keywords: Fuzzy Logic. Neural networks. Prediction.

### 1 Introdução

O tênis é um dos esportes mais populares do mundo, principalmente quando pensamos no universo dos esportes individuais. Com um tour anual composto por cerca de oitocentos torneios, espalhados por mais de setenta países [1, 2]. Os principais eventos atraem milhões de expectadores e distribuem premiações milionárias, pois esse esporte tem uma grande e fiel legião de fãs e seus principais atletas são alguns dos esportistas mais populares e bem-remunerados [3] do mundo.

http://dx.doi.org/10.5335/rbca.v8i2.5508

<sup>&</sup>lt;sup>1</sup> Instituto Federal de Sergipe, *campus* Aracaju, Av. Gentil Tavares, 1.166 – Aracaju (SE) - Brasil {mateus.fernandes@ifs.edu.br}

Com essa popularidade, o tênis movimenta ano após ano valores cada vez maiores com a comercialização dos ingressos, os contratos publicitários, o material esportivo e até mesmo com apostas, isto sem contar com as premiações oferecidas pelos torneios e com a valorização das imagens dos atletas. Paralelamente, a esse aumento no interesse comercial que permeia não apenas o tênis, mas os esportes profissionais em geral, há também uma presença cada vez mais marcante de metodologias científicas quantitativas aplicadas a análises. Esses métodos se tornaram imprescindíveis tanto para que atletas e treinadores analisem performance, estratégias, deficiências, pontos fortes [4] e até mesmo aspectos da preparação física e biomecânica [5], quanto para que organizadores, investidores e mídia disponham de informações importantes para seu planejamento e análises de viabilidade econômica.

Nesse contexto, o desenvolvimento de preditores de resultados é uma das vertentes de estudos, buscando gerar dados que podem ser interessantes não apenas para uso informativo ou como fonte de rendimentos em apostas, mas também para o planejamento do torneio e de sua cobertura televisiva. Prever partidas de ocorrência mais provável em rodadas futuras e/ou estimar seus tempos de duração pode, por exemplo, auxiliar na alocação de jogos atrativos nas principais quadras e nos melhores horários, permitir previsões de público e audiência e até mesmo embasar ações de merchandising [6, 7]. Na literatura, vários trabalhos são encontrados tratando da previsibilidade de resultados empregando as mais diversas abordagens, incluindo análises ponto a ponto no decorrer da partida e previsões de vencedores antes do início de cada jogo.

Os trabalhos de Clowes et al. [8] e de Klaasen e Magnus [6] são alguns dos que se basearam em análises ponto-a-ponto, focando não apenas na previsão antes da partida como também - e principalmente - no seu decorrer, simulando com base nas chances que o jogador que está servindo tem de vencer o próximo ponto. Knottenbelt et al. [9] também apresentam um preditor para partidas com análise a cada ponto, no entanto acrescentando informações de desempenho dos jogadores em questão contra um adversário pregresso em comum, de forma a eliminar o viés existente nas estatísticas de servico de que jogadores mais fortes, por normalmente avançarem mais frequentemente às rodadas finais dos torneios, têm em média adversários também mais fortes. Clarke e Dyte [10] ajustaram um modelo de regressão logística para cálculo da probabilidade de se vencer um set com base nas diferenças de pontuação no ranking entre os jogadores. Esse modelo foi empregado para a previsão de partidas e também para a simulação de torneios.

Os trabalhos supracitados baseiam-se na hipótese de que os pontos (ou sets) são eventos independentes e identicamente distribuídos (i.i.d.), ou seja, que resultados anteriores não influenciam nos resultados seguintes. No entanto, o trabalho de Klaasen e Magnus [11] discute a validade dessa hipótese, concluindo que vencer o ponto anterior tem uma influência positiva na disputa do ponto atual e que em pontos disputados sob condições de maior responsabilidade, os sacadores são negativamente afetados, hipóteses que são mais verossímeis.

Na abordagem de del Corral e Prieto-Rodríguez [12], que se consiste em uma previsão do vencedor em jogos do Grand Slam, sem se ater ao placar, foram analisadas as influências de variáveis, como o tipo de superfície e as características físicas dos jogadores nos resultados de partidas, além do ranking de cada tenista. McHale e Morton [7] realizam as predições empregando um modelo de Bradley-Terry (baseado em comparações entre pares de dados) ajustado a partir de resultados prévios e na superfície de disputa. Já Scheibehenne e Bröder [13] mostram que é possível obter bons índices de acertos apenas com o reconhecimento de nomes de jogadores por um público não necessariamente especializado.

No presente trabalho, tal como nos artigos citados no parágrafo anterior, também são realizadas previsões de vencedores por partida, antes de seu início e sem levar em conta quaisquer eventos que possam acontecer no seu decorrer. No entanto, diferentemente dos trabalhos correlatos que se baseiam em métodos estatísticos para as previsões, esse estudo propõe uma abordagem baseada em técnicas de inteligência computacional. Dois diferentes preditores são desenvolvidos: o primeiro, empregando um sistema de inferência Fuzzy baseado em pertinências e regras que tentam mimetizar o conhecimento de um especialista, e, o segundo, empregando uma rede neural artificial (RNA) e explorando suas capacidades de aprendizado e extração de características a partir de conjuntos de treinamento compostos por históricos de partidas. Para um melhor aproveitamento das qualidades dessas técnicas, é feita uma análise prévia focada em obter correlações entre alguns fatores de influência nas partidas que são capazes de ser quantificados e a crença em qual dos jogadores sairá vencedor em uma partida específica.

A metodologia empregada e a aquisição do conjunto de dados utilizado são apresentadas na seção 2, que é seguida por uma análise dos fatores de influência nos resultados de partidas, mostrada na seção 3. Após, a seção 4 contém os detalhes da implementação do preditor Fuzzy e a seção 5, do preditor por redes neurais. Na seção 6 subsequente, os resultados dos preditores para as partidas são apresentados e seus desempenhos analisados. Finalmente, a seção 7 traz as considerações finais.

### 2 Aquisição de dados e metodologia

Para emprego dos métodos quantitativos propostos, um banco de dados foi levantado, contando com estatísticas de 220 jogadores em atividade no circuito profissional masculino em 2014 e 2015. Esses dados têm como fonte os números fornecidos pela Associação dos Tenistas Profissionais (ATP) em seu site oficial [1] e são compostos por:

- Número de títulos acumulado na carreira:
- Fração de jogos vencidos em toda a carreira do jogador;
- Fração de vitórias nas diferentes superfícies grama, saibro e piso sintético para cada jogador;
- Fração de jogos de *Grand Slam* vencidos em toda a carreira do jogador.

As estatísticas computam apenas resultados de partidas disputadas nas chaves principais dos torneios de classes mais elevadas, os ATPs e Grand Slam [1, 2], ou seja, são desconsiderados os resultados em torneios de nível Challenger e Future, bem como em qualificatórios, o que é uma tentativa de padronizar os níveis de dificuldade dos jogos, mantendo isonomia nos dados. Um banco de dados adicional, com os resultados de todas as partidas desses níveis, disputadas nas temporadas mais recentes, foi obtido em [14]. Esse conjunto de dados contém ainda a posição e a pontuação no ranking de entradas da ATP de cada um dos atletas ao início do torneio em questão. Todos esses dados empregados foram atualizados em diferentes momentos para estarem condizentes com as previsões requeridas.

De posse das informações descritas, o estudo inicia com uma avaliação cujo propósito é descobrir quais são os fatores de maior contribuição para as previsões mais eficientes de resultados. Após identificadas as principais relações, elas são tomadas como base para o desenvolvimento, de forma independente, dos dois preditores propostos. Para maximizar o aproveitamento das capacidades preditivas das técnicas de inteligência computacional empregadas, a análise dos fatores quantitativos é a responsável pela definição de quais serão as melhores variáveis a serem usadas como entradas para os preditores.

Durante o desenvolvimento dos preditores propriamente ditos, cujos detalhes de implementação são discutidos nas seções 4 e 5, o conjunto de dados relativo aos torneios disputados em 2014 são empregados para treinamento e ajustes. Os dados dos torneios de 2015 constituem o conjunto de teste, apresentado apenas às versões finais dos preditores e permitindo que seja feito um estudo comparativo do seu desempenho na tentativa de acertar os vencedores de partidas e torneios.

### 3 Análise de fatores de influência nas previsões

## 3.1 Influência do ranking

O ranking de entradas da ATP classifica os tenistas com base em suas pontuações acumuladas nos torneios disputados nas últimas 52 semanas, com o propósito de definir as entradas e o chaveamento dos torneios. Neste trabalho, seu uso é proposto como forma de avaliação do desempenho recente dos jogadores.

Ao se tratar do uso de informação do ranking para as predições, uma observação constatada é que, com base nos históricos de partidas da ATP, pode-se perceber uma forte tendência de que a dificuldade de um tenista vencer um adversário parece aumentar a passos cada vez mais largos à medida em que o ranking desses adversários se aproxima do pináculo. Ou seja, é muito mais comum uma vitória do 120º classificado sobre o 101º do que uma vitória do 20º sobre o líder, embora a diferenca de posições seja a mesma. Logo, a relação de força que parece existir não está ligada somente à posição no ranking, o que induz a pensar em uma relação matemática para sua modelagem. Para compreender um pouco melhor essa tendência, são plotadas no gráfico da Figura 1 curvas de pontuação no ranking em função da posição para cinco diferentes datas, entre os anos 2012 e 2015 (período sem mudanças de critérios na pontuação dos torneios).

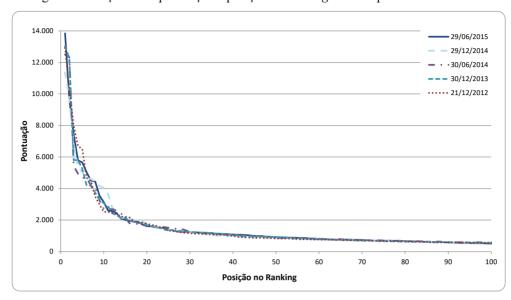


Figura 1: Relação entre pontuação e posição no ranking da ATP para diferentes datas

Quando essas curvas são analisadas, a relação obtida parece se assemelhar à tendência descrita, com diferenças de pontuação cada vez mais marcantes à medida que nos aproximamos do topo do *ranking*. Isso faz sentido pela maneira como a lista classificatória é concebida, considerando que, assim como a dificuldade dos adversários tende a crescer aceleradamente, as pontuações concedidas a cada rodada avançada em um torneio também crescem de forma geométrica.

Utilizando curvas para diversas datas foi possível até mesmo modelar esse comportamento, tal qual feito por Clarke e Dyte [10], relacionando a pontuação à posição no *ranking*. Para o presente caso, o melhor ajuste encontrado foi baseado em uma equação de potência, cujos parâmetros foram ajustados com minimização dos erros quadráticos resultando em (1):

$$Pontuacão = 18157 \cdot Posicão^{-0.779}$$
 (1)

À luz dessas informações, foi proposto como fator de influência no desempenho esperado dos jogadores, o emprego simples de sua pontuação no *ranking* normalizada em função da pontuação corrente do líder.

## 3.2 Desempenho ao longo da carreira

Do banco de dados obtido, uma das possíveis formas de quantificar o desempenho de jogadores ao longo de suas carreiras é por meio das frações de jogos vencidos (tanto em números gerais como em pisos específicos, conforme disponibilizados pela ATP); entretanto, essas frações nem sempre serão confiáveis, devido, principalmente, à diferença entre o número de jogos disputados na carreira de cada tenista. Ilustrando com um caso real, o jovem tenista Jiri Vesely, no momento de um dos levantamentos de dados para este estudo, havia disputado apenas três jogos sobre grama em torneios de primeira linha e vencido dois, resultando em um bom aproveitamento de 0,667. Porém, na prática esse valor não deveria ser mais significativo que a fração 0,656 obtida com 59 vitórias e 31 derrotas pelo muito mais experiente Ivo Karlovic.

Por razões como essa é que os modelos de previsão devem se basear também em outros fatores para constituir um histórico de longo prazo. Por essa razão, propõe-se aqui o acréscimo da informação do número de títulos em uma tentativa de se retratar mais fielmente a carreira de cada jogador. A aplicação dessa variável é proposta, considerando apenas seu número absoluto, sem uso de fatores de ponderação por sua importância. A quantificação para uso como fator de desempenho nos preditores é feita simplesmente com uma normalização que tem como referência o maior número de títulos dentre os tenistas em atividade, no caso Roger Federer.

## 3.3 Superfície dos jogos

Embora tenha surgido como um esporte disputado sobre a grama, o tênis conta hoje com três diferentes classes de piso: os duros (que por sua vez englobam uma variedade de pisos sintéticos), o saibro e a própria grama, atualmente adotada em um reduzido número de torneios. Cada um destes tipos de piso – por sua influência na velocidade do jogo, no quique da bola e na movimentação em quadra – apresenta particularidades nas exigências físicas, técnicas e táticas, requerendo grande adaptabilidade por parte dos jogadores e, muitas vezes, resultando em diferenças significativas de desempenho.

A fração de vitórias em uma superfície específica é aqui considerada em função desse fato, sendo um fator importante para auxílio nas previsões, uma vez que as diferentes superfícies impõem diferentes características ao jogo. Por exemplo, sobre a grama, considerado o piso mais rápido, os jogadores reconhecidamente detentores de um bom serviço e de uma maior capacidade de jogar agressivamente, incluindo subidas à rede e uso de voleios para encurtar os pontos, geralmente conseguem ter seu melhor rendimento. Já no saibro, superfície mais lenta, normalmente adaptam-se melhor os jogadores com boas capacidades defensivas e de movimentação eficiente no fundo de quadra, o que tem uma correlação com o desempenho nos pontos com maior número de trocas de bola. Isto é evidenciado quando comparamos, por exemplo, estilos de jogo e resultados nas duas superfícies dos maiores campeões em atividade, Roger Federer e Rafael Nadal, o primeiro com um estilo mais ofensivo e o maior vencedor da era profissional sobre quadras de grama, enquanto que o segundo, com sua eficiência defensiva, é o maior campeão sobre quadras de saibro.

O estudo de Clarke e Dyte [10] chega a comparar a preferência dos jogadores por determinado tipo de piso à vantagem da partida jogada em casa observada nos esportes coletivos, como o futebol ou o basquete. É importante, inclusive, para a confiança do jogador, uma vez que, no tênis, jogar em seu país de origem, muitas vezes, não implica em uma diferença significativa de desempenho, conforme apontado por Holder e Nevill [15].

Já o trabalho de Barnett e Pollard [16] analisa o desempenho de jogadores nas diferentes superfícies, mostrando que aqueles com melhor desempenho em quadras de grama dificilmente têm como segunda melhor superfície a quadra de saibro (e vice-versa), sendo as quadras duras, como a DecoTurf usada no US Open e a Rebound Ace usada no Australian Open, um meio-termo entre elas.

Uma análise no banco de dados levantado para este trabalho leva a uma conclusão similar, quando se quantificam as correlações entre os desempenhos nas diferentes superficies com o emprego do coeficiente de Pearson, também conhecido como coeficiente de correlação produto-momento. Esse coeficiente representa uma medida do grau de uma relação linear entre pares de dados e é calculado por (2):

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2\right]} \left[\sum_{i=1}^{n} (y_i - \overline{y})^2\right]}$$
(2)

em que x e y são os vetores de dados, ambos com n valores. Valores de r próximos da unidade representam relações lineares fortes, enquanto que valores nulos mostram ausência de uma relação linear [17]. Para o grupo estudado, a correlação entre os desempenhos calculada para o par saibro-dura é de 0,688 e para grama-dura é de 0,719, indicando claramente uma correlação mais forte do que a obtida para o par grama-saibro, de 0,528.

## 3.4 Jogos de Grand Slam

Outra variável de interesse é a fração de jogos vencidos em torneios do *Grand Slam*, classe composta pelos mais tradicionais e prestigiosos torneios do circuito: Australian Open, Roland Garros, Wimbledon e US Open. Esses torneios são os únicos a contar com 128 tenistas em sua chave principal e, no masculino, a ter suas partidas disputadas em melhor de cinco *sets*. Por isso, as premiações e pontuações concedidas aos vitoriosos são também mais generosas. Observa-se, nesse caso, uma mudança de comportamento nas estatísticas, o que pode estar relacionado aos aspectos mental e físico, pois são partidas mais longas, fazendo parte dos maiores eventos. A pressão sobre um grande jogador, por exemplo, é grande desde as rodadas iniciais, quando há a "obrigação" de vencer adversários geralmente menos expressivos. Já os jogadores menos gabaritados, normalmente, tendem a

enfrentar os grandes nomes desde a primeira rodada, além de, muitas vezes, não estarem acostumados a lidar com a pressão dos grandes torneios.

Comprovação dessa diferença pode ser observada na comparação entre as Figuras 2 e 3. A primeira mostra as distribuições de frequência para os desempenhos do grupo de jogadores analisado nas três diferentes classes de pisos para os torneios ATP, em que se percebe que elas se aproximam de distribuições gaussianas simétricas com média 0,5. Entretanto, a segunda curva, com os desempenhos nas partidas do *Grand Slam*, apresenta uma feição pouco diferente, com seu pico deslocado em direção aos valores mais baixos.

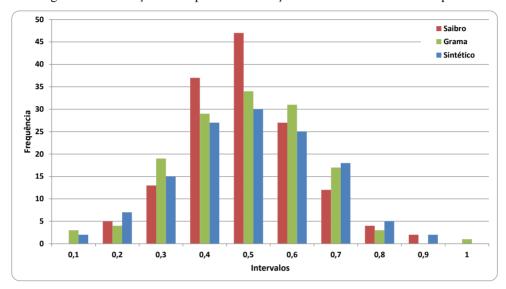


Figura 2: Distribuição de frequências das frações de vitórias nos diferentes pisos

Fonte: elaborado pelo autor com dados da pesquisa.

Esse deslocamento do pico pode ser explicado pela tendência de que as vitórias de jogadores com *ranking* mais baixo tornam-se mais escassas, ou seja, um menor número de jogadores tende a concentrar o maior sucesso nesses torneios. Essa maior taxa de confirmação de favoritismo em *Grand Slam* é também observada na modelagem de Clarke e Dyte [10], que ressaltam que os favoritos, sendo mais propensos a vencer *sets*, terão derrotas em partidas de melhor de cinco *sets* mais dificultadas do que em partidas de melhor de três *sets*. Essa constatação vai ao encontro dos fatos, dado que nos últimos anos a maior parte desses grandes títulos (quarenta dentre os 48 disputados entre 2004 e 2015) ficou nas mãos de apenas três jogadores: o suíço Roger Federer, o espanhol Rafal Nadal e o sérvio Novak Djokovic, que, por tal feito, são reconhecidos como alguns dos maiores nomes da história nesse esporte.

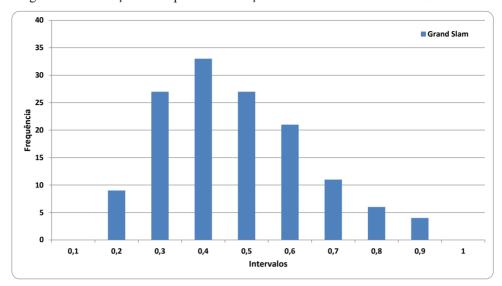


Figura 3: Distribuição de frequência das frações de vitórias nos torneios de Grand Slam

## 4 Desenvolvimento do preditor Fuzzy

Desde o trabalho seminal publicado por Zadeh nos anos 1960 [18], a lógica Fuzzy é empregada em uma vasta gama de problemas, sendo que os sistemas de inferência constituem uma de suas aplicações mais difundidas. Introduzidos por Mamdani, sistemas desse tipo são regidos pela forma de raciocínio aproximado, conhecida como *modus ponens* generalizado, baseando-se em variáveis linguísticas e regras de implicação do tipo SE–ENTÃO para gerar um raciocínio aproximado, característico dos sistemas Fuzzy [19] [20]. A solução proposta neste trabalho baseia-se em um sistema de inferência de Sugeno de ordem zero [21], em que o consequente das regras de implicação é uma constante.

No preditor Fuzzy desenvolvido, são utilizados como dados de entrada três variáveis, sendo cada uma delas introduzida na forma de uma diferença entre os respectivos valores para os jogadores da partida em questão. A primeira diferença é entre os valores correntes da pontuação no *ranking*, normalizados em relação à pontuação do líder naquele mesmo momento. A segunda variável corresponde à diferença entre os históricos dos tenistas, sendo os valores para cada um deles quantificado por meio da média aritmética entre a fração de jogos (nível ATP e *Grand Slam*) vencidos ao longo de toda a carreira e o coeficiente de títulos calculado conforme descrito na seção 3.2. A terceira variável consiste na diferença entre as frações de jogos vencidos ao longo da carreira dos jogadores na superfície específica do torneio em questão. Nesse caso, foi feita a opção de não se incluir as informações do desempenho em *Grand Slam* na modelagem do preditor.

Para cada partida a ser prevista, essas três entradas são calculadas e então fuzzificadas a partir da divisão em quatro categorias de valores – grande negativa, pequena negativa, pequena positiva e grande positiva – por meio de funções de pertinência triangulares, que são definidas de modo genérico por (3):

$$trimf(x;a,b,c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases}$$
 (3)

e com a configuração final mostrada na Figura 4, sendo igual para as três variáveis propostas. Com as pertinências assim definidas, são calculados os graus de compatibilidade que fornecem a crença nos antecedentes de cada uma das regras. As funções triangulares são escolhidas por sua eficiência e simplicidade matemática, implicando em um custo computacional reduzido. Não foram observadas melhorias de desempenho nas

previsões ao serem trocadas as funções triangulares por gaussianas, nem tampouco com otimização no número de classes no ajuste fino dos parâmetros de definição das mesmas.

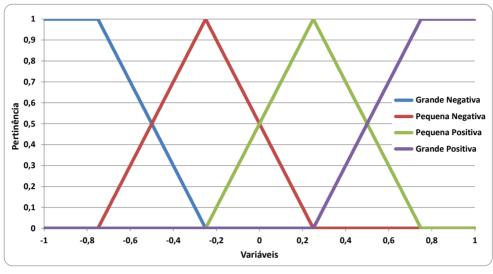


Figura 4: Pertinências utilizadas no preditor Fuzzy

Fonte: elaborado pelo autor com dados da pesquisa.

Para a geração do conjunto de regras que definem o sistema de inferência, a experiência humana é a fonte de informação primária, sendo que a lógica Fuzzy mostra aqui a sua principal vocação, permitindo que se possa expressar matematicamente o conhecimento que comumente lidamos na forma linguística. Dessa forma, instituindo os dois resultados possíveis – vitória do jogador 1 ou vitória do jogador 2 – é construída uma base de regras para análise das três variáveis de entrada. Uma amostra de algumas dessas regras é mostrada na Tabela 1. Os operadores lógicos E são implementados pela função *mínimo* conforme (4):

$$\mu_C(x) = \min(\mu_A(x), \mu_B(x)) \tag{4}$$

em que  $\mu_A$  e  $\mu_B$  são as funções de pertinência escolhidas.

**ARanking** ΔHistórico **ASuperfície** Resultado SE Grande Positiva  $\overline{E}$ Grande Positiva **ENTÃO** Vitória P1  $\boldsymbol{E}$ Pequena Positiva SEPequena Positiva EPequena Negativa *ENTÃO* Vitória P1 SEGrande Negativa  $\boldsymbol{E}$ Pequena Positiva  $\boldsymbol{E}$ Grande Positiva *ENTÃO* Vitória P1 Grande Positiva SE $\boldsymbol{E}$ Pequena Negativa  $\boldsymbol{E}$ Grande Negativa **ENTÃO** Vitória P2 SEPequena Negativa  $\boldsymbol{E}$ Pequena Negativa  $\boldsymbol{E}$ Pequena Positiva *ENTÃO* Vitória P2 **ENTÃO** SEGrande Negativa Grande Negativa Vitória P2

Tabela 1: Excerto da base de regras do sistema de inferência projetado

Fonte: elaborado pelo autor com dados da pesquisa.

O procedimento descrito até este ponto fornece os fatores de ponderação para cada um dos resultados. Aqui, diferentemente de outros sistemas de inferência, não há interesse em um valor único de saída, o que requereria uma defuzzificação, e, sim, nos valores das crenças em cada um desses resultados. Isto leva à adoção dos próprios fatores de ponderação, depois de devida normalização, como as crenças desejadas, expressas, enfim, na forma de percentuais para cada jogador. Credita-se então a vitória ao jogador cuja saída correspondente apresente o valor mais alto. O sistema de inferência para previsão, na forma como descrito, é

ilustrado pela Figura 5. Os resultados obtidos por este modelo de preditor, acompanhados de sua devida análise e comparação com outros modelos, são apresentados na seção 6.

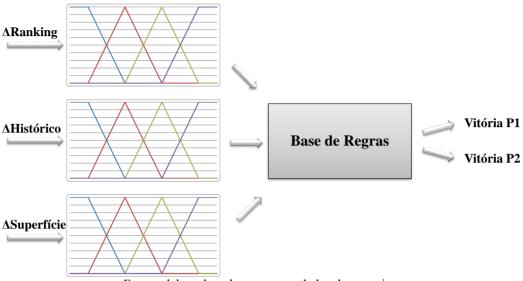


Figura 5: Modelo do sistema de inferência utilizado como preditor

Fonte: elaborado pelo autor com dados da pesquisa.

### 5 Desenvolvimento do preditor por redes neurais

Uma rede neural artificial, no paradigma original inspirado por redes neurais biológicas, é constituída por um conjunto de unidades de processamento (também chamados de neurônios ou nodos) destinado a fornecer um valor de saída dentro de um determinado intervalo com base na soma ponderada de suas entradas e subsequente aplicação de uma função de ativação. Diversos são os possíveis tipos de conexão entre essas unidades para a formação das redes [22], sendo o modelo mais difundido o MLP (sigla do inglês para Perceptron de Múltiplas Camadas), uma rede com propagação de sinal direta, cujos neurônios são dispostos em camadas sequenciais e as saídas de todos os neurônios de cada camada são conectadas às entradas da camada seguinte. O "conhecimento" da rede é armazenado justamente nos pesos associados a cada uma dessas conexões, com seu "aprendizado" sendo feito usualmente em processos iterativos de ajustes destes pesos com base em exemplos conhecidos.

Uma rede MLP com uma camada intermediária e uma camada de saída já é capaz de resolver alguns problemas não lineares e de aproximar funções contínuas, o acréscimo de mais uma camada intermediária a capacita a implementar qualquer função, linearmente separável ou não, conforme demonstrado por Cybenko [23]. O número de nodos em cada camada é o principal responsável pela convergência no treinamento e pela precisão dos resultados [22]. Cada um dos neurônios da RNA é portador de uma função de ativação, responsável pelo cálculo do valor da saída do nodo em função da soma ponderada de suas entradas. As ativações mais comuns são as baseadas em funções sigmoides (em forma de S), devido ao seu balanceamento entre comportamento linear e não linear, sendo ainda funções diferenciáveis em todos os pontos [24]. As funções sigmoides mais comumente empregadas na literatura, e também escolhidas para uso neste trabalho, são a logística, expressa por (5):

$$f(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

e a tangente hiperbólica, dada por (6):

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$
 (6)

Essas funções têm ainda como propriedade a compressão dos valores de entrada, sendo que grandes valores positivos se aproximam assintoticamente de 1 e os grandes valores negativos tendem a 0. Outros exemplos de funções de ativação podem ser encontrados em [24].

No presente trabalho, uma rede desse tipo é então treinada para, com base em um conjunto de dados de partidas de vencedor já conhecido e com os dados de histórico e desempenho prévios dos jogadores, utilizar sua capacidade de generalização para prever o vencedor em novas partidas, quando dadas as novas entradas. A arquitetura que se mostrou mais adequada à resolução desse problema, determinada empiricamente, é a mostrada na Figura 6. Essa RNA é formada por uma camada de entrada com quatro neurônios e funções de ativação tangente hiperbólica, uma camada oculta com quatro neurônios e funções de ativação logística e uma camada de saída com dois neurônios e funções de ativação tangente hiperbólica. Todas as conexões entre nodos são ponderadas por um peso específico: vetor  $W_i$  para a camada de entrada,  $W_h$  para a intermediária e  $W_o$  para a de saída.

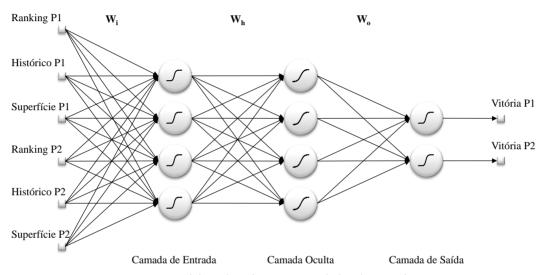


Figura 6: Modelo da rede neural utilizada para a predição

Fonte: elaborado pelo autor com dados da pesquisa.

Como variáveis de entrada para a rede neural, no modelo final proposto, são empregados os mesmos três medidores de desempenho nos quais o preditor Fuzzy se baseia: coeficientes de pontuação atual no *ranking*, de histórico do atleta (composição de títulos e fração de jogos vencidos) e fração de jogos vencidos na superfície do torneio em questão. Porém, aqui são empregados os valores individuais para os dois atletas que se enfrentam em uma partida específica, totalizando seis entradas. As variáveis de saída, por sua vez, são duas, cada uma representando a vitória de um dos jogadores em valores binários. Assim, para vitória do primeiro jogador, esperamos que a saída correspondente tenha valor unitário e que a outra saída tenha valor zero, por exemplo. A

Tabela 2 ilustra alguns dos dados utilizados como treinamento. Também foram avaliadas redes contando com a adição dos dados de desempenho em *Grand Slam*, perfazendo um total de oito entradas.

**Torneio Partida** Placar **Rk.** 1 Hist. 1 Sup. 1 Rk. 2 Hist. 2 Sup. 2 **P1** P2 1-6 6-7 2-6 Wimbledon'14 Murray-Dimitrov 0,374 0,757 0,830 0,208 0,470 0,622 0 1 US Open'14 Cilic-Federer 6-3 6-4 6-4 0,144 0,611 0,670 0,587 0,887 0,829 1 0 Paris'14 Djokovic-Raonic 6-2 6-3 0,817 0,826 0,828 0,279 0,546 0,710 1 0 0,509 Xangai'14 Nadal-Lopez 3-6 6-7 0,600 0.870 0,776 0,124 0,435 0 1 Toronto'14 Ferrer-Dodig 1-6 6-3 6-3 0,290 0,682 0,637 0,057 0,312 0,483 1 0

Tabela 2: Excerto do conjunto de dados de treinamento para a rede neural

Fonte: elaborado pelo autor com dados da pesquisa.

O aprendizado da RNA nesse trabalho deu-se pelo mais popular dos métodos de treinamento, o backpropagation. Nesse método, os pesos das conexões entre os nodos da rede são iniciados com valores aleatórios. Em seguida, são apresentados à rede (em ordem também aleatória) conjuntos de valores de entrada que resultam em uma saída já conhecida. Para cada um desses conjuntos, a saída da rede com os pesos atuais é computada, naquela que é chamada de fase forward do treinamento. A saída obtida então é comparada ao valor correto desejado para cálculo da diferenca entre ambas. Esse erro é propagado de forma reversa pela rede (fase backward, daí o nome do algoritmo). O produto do erro de cada saída por uma "taxa de aprendizado" constante é subtraído dos pesos das conexões do respectivo nodo da última camada. O erro de cada nodo das camadas anteriores é calculado usando os erros dos nodos da camada seguinte conectados a ele, ponderados pelos pesos das conexões entre eles [22]. O procedimento repete-se com novos pares de vetores entrada/saída, que são apresentados à rede até que um critério de parada seja atingido; que o erro quadrático médio torne-se menor do que um limite pré-estabelecido, que um número máximo de iterações seja efetuado ou que o erro fique estagnado entre as iterações.

#### 6 Resultados

A análise do desempenho dos preditores apresentados baseou-se no banco de resultados de partidas das últimas temporadas disponível em [14]. Para a constituição do conjunto de dados de treinamento para a rede neural e também para o ajuste fino do preditor Fuzzy foram utilizadas as informações de 1.744 jogos disputados em torneios das categorias ATP 250, ATP 500, Masters 1000 e ATP Finals, somados aos 508 jogos nos torneios de Grand Slam, todos no ano de 2014. Essas partidas envolveram mais de duzentos diferentes jogadores.

Os testes dos preditores e as estatísticas das suas respostas basearam-se em um conjunto que conta com 1.109 jogos em torneios ATP somados às 381 partidas disputadas nos torneios Australian Open, Roland Garros e Wimbledon, sendo todo esse conjunto relativo à temporada de 2015. Novamente foram incluídos mais de duzentos diferentes tenistas nesse grupo de partidas.

Do conjunto de partidas realizadas em 2014, o levantamento mostra que o percentual de jogos vencidos pelos jogadores de melhor ranking foi próximo dos 67% nos torneios ATP e dos 74% nos Grand Slam, o que é bastante condizente com as médias observadas nos anos anteriores, conforme exposto na Tabela 3. Essa comparação é importante para verificação de que as análises não estão sendo baseadas em dados de eventos atípicos. A mesma tabela mostra ainda, para efeito comparativo, os percentuais de acertos obtidos a partir de dados de cinco dos principais sites de apostas para jogos de tênis, conforme números também compilados em [14]. Esses valores representam a fração de jogos que foram vencidos pelos jogadores que eram considerados os mais cotados para cada jogo antes de seu início, o que representa um bom benchmark para preditores, por serem números que representam a confiança de apostadores que em sua maioria possuem conhecimento prévio sobre os jogadores (performance passada, forma atual e eventuais particularidades) e as características do torneio, podendo analisar jogo a jogo de forma automatizada ou não.

Tabela 3: Desempenho de previsões por ranking e por média de apostas – últimos 5 anos

	ATPs		Grand Slam	
Ano	% Acertos Ranking	% Acertos Apostas	% Acertos <i>Ranking</i>	% Acertos Apostas
2010	64,72%	67,12%	74,95%	78,59%
2011	66,14%	69,40%	75,00%	78,22%
2012	66,07%	68,92%	74,85%	77,86%
2013	64,00%	66,98%	75,31%	77,97%
2014	66,85%	67,80%	74,07%	75,98%

Fonte: elaborado pelo autor com dados da pesquisa.

Resultados imprevisíveis, que poderiam ser outliers no conjunto tomado, não foram retirados do conjunto de treinamento. Essa decisão tem o propósito de permitir que os preditores possam justamente tentar extrair algum padrão para os resultados que seriam normalmente considerados como grandes zebras, segundo o jargão esportivo.

## 6.1 Previsão de partidas

Conforme citado anteriormente, as previsões neste trabalho são feitas para os vencedores das partidas, sem que haja preocupação com eventos no seu decorrer ou até mesmo com o placar final. Sendo assim, para ambos os preditores as informações necessárias são as entradas de cada partida a ser analisada, conforme detalhamento nas seções 4 e 5, e o vencedor da partida, para comparação com as respostas obtidas.

A Tabela 4 apresenta os resultados dos preditores propostos para o conjunto de jogos nível ATP usados como conjunto de teste, sendo mostrados com quebra por classes de superfícies. Vale ressaltar que, embora sejam apresentados os resultados com essa segmentação, não foram empregados modelos específicos por piso. O sistema de inferência Fuzzy é único para todas as previsões e a rede neural é treinada com todo o conjunto de partidas de torneios ATP sem distinções, sendo que a essa única Rede são apresentados todos os dados de entrada para teste. Os percentuais de acerto são comparados àqueles obtidos pela mera comparação entre os rankings dos jogadores à época do torneio e também das cotações de apostas como citado.

Superfície	Nº de Partidas	% Acertos Ranking	% Acertos Apostas	% Acertos Rede neural	% Acertos Fuzzy
Dura	474	62,87%	69,62%	71,66%	65,82%
Saibro	483	68,74%	70,73%	77,02%	70,60%
Grama	152	55,92%	66,77%	72,15%	62,50%
Total	1.109	64,47%	69,75%	74,06%	67,45%

Tabela 4: Performance dos preditores – torneios nível ATP

Fonte: elaborado pelo autor com dados da pesquisa.

Desses resultados, pôde-se observar que a taxa de acertos do preditor Fuzzy representa um ganho em relação à previsão por ranking, porém ainda tem desempenho inferior ao dos apostadores. Já a rede neural conseguiu uma taxa de acerto superior a todos os demais, com uma diferença bastante significativa, o que quer dizer que ela foi capaz de extrair do conjunto de treinamento características relevantes e de quantificá-las em seu modelo de forma eficiente.

A Tabela 5 apresenta os resultados para as previsões de partidas dos torneios Grand Slam, com as mesmas comparações. Nesse caso, o preditor Fuzzy seguiu a mesma modelagem e a mesma base de regras daquele utilizado para os torneios ATP, enquanto que a rede neural empregada, embora de configuração idêntica à anterior, foi treinada apenas pelo conjunto composto pelas 508 de partidas do Grand Slam disputadas em 2014.

Grand Slam	Nº de Partidas	% Acertos <i>Ranking</i>	% Acertos Apostas	% Acertos Rede neural	% Acertos Fuzzy
Australian Open	127	74,80%	78,74%	80,71%	77,95%
<b>Roland Garros</b>	127	71,65%	78,15%	77,56%	76,38%
Wimbledon	127	75,59%	75,59%	71,65%	74,02%
Total	381	74,02%	77,49%	76,64%	76,12%

Tabela 5: Performance dos preditores – torneios nível *Grand Slam* 

Fonte: elaborado pelo autor com dados da pesquisa.

Para esse modelo, com o mesmo conjunto de jogos também foram realizados testes, incluindo uma nova variável de entrada: a fração de jogos de Grand Slam vencidos. No entanto, a adição dessa variável não resultou em melhoria de qualidade nas previsões, sendo os resultados mostrados baseados em redes com a mesma configuração apresentada na Figura 6.

Para essa classe de torneios, como os índices de vitórias dos melhores classificados já é significativamente superior, é de se esperar que as margens de melhoria sejam menores. Observando-se a Tabela 5, é exatamente o que pode ser constatado. Aqui, os dois preditores projetados melhoraram os números obtidos pela previsão por ranking, mas ambos ficaram pouco abaixo do desempenho dos apostadores. Nesse caso, ficam mais claras as limitações das previsões automáticas, já que nunca poderão contemplar todos os aspectos quantitativos e qualitativos que um preditor humano (como os apostadores) poderia levar em consideração, por

exemplo, a influência da torcida, a fadiga gerada em rodadas anteriores, as alterações momentâneas nas condições físicas e emocionais, motivações ou pressões extras, etc.

### 6.2 Previsão de torneios

Após o ajuste do preditores para partidas, foram estudadas extensões para previsões de torneios, dada sua chave inicial. Foram realizadas simulações baseadas apenas no preditor Fuzzy, que é capaz de fornecer crenças nas vitórias de cada um dos jogadores que se enfrentam em uma partida. A rede neural, pelas características da função de ativação empregada na camada de saída não é apropriada a realizar essas simulações por não haver uma gradação nas respostas, que tendem aos valores de 0 para o perdedor ou 1 para o vencedor. Para essas previsões de torneios - crenças percentuais no título para cada competidor - foi adotada uma simulação baseada na propriedade tratada no estudo das probabilidades, a Lei dos Grandes Números (7):

$$p = \lim_{n \to \infty} \frac{S_n}{n} \, \cdot \tag{7}$$

Essa lei enuncia que a probabilidade de um evento é a razão entre o número de ocorrências  $S_n$  e o número de ensaios n quando esse tende a infinito [25]. Dessa forma, de posse da chave inicial de determinado torneio, o sistema de inferência foi utilizado para previsões probabilísticas de cada partida. Como o preditor é capaz de gerar crenças na vitória de cada tenista, em cada partida simulada o resultado é determinado por meio do sorteio de um número aleatório, uniformemente distribuído entre 0 e 1 e sua comparação com essas crenças. Por exemplo, na partida entre Andy Murray e David Ferrer, nas quartas de final de Roland Garros 2015, o preditor calculou uma crença de 0,664 na vitória do primeiro, com os outros 0,336 para o segundo; logo, na simulação da partida, caso o número aleatório sorteado esteja entre 0 e 0,664 a vitória será computada para Murray e caso esteja entre 0,665 e 1 para Ferrer. Com base nos resultados simulados para cada rodada, são então montados os confrontos para a próxima rodada de acordo com o emparelhamento definido no sorteio que realmente ocorreu para o torneio em questão, até que se chegue ao campeão para essa iteração da simulação. Para a obtenção de probabilidades mais confiáveis baseadas na Lei dos Grandes Números, foram realizadas dezenas de milhares de simulações para cada torneio a partir de seu chaveamento inicial real, e sempre foram registrados para cada jogador o número de ocorrências de título e de presença nas quartas de final. Outros conjuntos de simulações foram feitos registrando as chances de título partindo das quartas de final que realmente ocorreram nos torneios (e não nas simuladas).

Os resultados das simulações realizadas para os torneios de Roland Garros 2015 e Wimbledon 2015 são apresentados nas Tabelas 6 e 7, respectivamente. O primeiro par de colunas mostra os oito jogares que, dentre os 128 que iniciaram o torneio, mais vezes apareceram nas quartas de final nas simulações, e seus respectivos percentuais de ocorrência; no segundo par de colunas, constam os valores para a ocorrência de títulos e, no terceiro par, os percentuais de títulos para os oito jogadores que efetivamente fizeram as quartas de final nesses torneios.

Tabela 6: Resultados da simulação do torneio de Roland Garros 2015

Jogador	% Quartas	Jogador	% Título	Jogador	% Título (a partir das quartas)
Nadal R.	69,47	Djokovic N.	31,17	Federer R.	44,51
Djokovic N.	65,83	Federer R.	30,64	Djokovic N.	35,67
Federer R.	63,63	Nadal R.	16,25	Nadal R.	8,84
Murray A.	62,00	Murray A.	9,72	Murray A.	7,66
Ferrer D.	52,30	Ferrer D.	2,91	Nishikori K.	1,07
Wawrinka S.	49,03	Nishikori K.	1,64	Ferrer D.	1,04
Nishikori K.	43,80	Wawrinka S.	1,46	Tsonga J.W.	0,70
Berdych T.	40,17	Berdych T.	1,43	Wawrinka S.	0,51

Fonte: elaborado pelo autor com dados da pesquisa.

Tabela 7: Resultados da simulação do torneio de Wimbledon 2015

Jogador	% Quartas	Jogador	% Título	Jogador	% Título (a partir das quartas)
Djokovic N.	66,33	Djokovic N.	33,90	Federer R.	43,58
Federer R.	66,00	Federer R.	33,11	Djokovic N.	40,89
Nadal R.	65,37	Murray A.	11,94	Murray A.	13,22
Murray A.	57,10	Nadal R.	7,89	Gasquet R.	0,80
Berdych T.	55,37	Berdych T.	1,79	Cilic M.	0,64
Wawrinka S.	39,37	Hewitt L.	1,61	Wawrinka S.	0,54
Nishikori K.	39,27	Cilic M.	1,32	Simon G.	0,25
Cilic M.	33,73	Tsonga J.W.	1,10	Pospisil V.	0,08

As Figuras 7 e 8 exibem os resultados do que de fato ocorreu nas três últimas rodadas.

Figura 7: Resultados do torneio de Roland Garros 2015, a partir das quartas de final



Fonte: elaborado pelo autor com dados da pesquisa.

Djokovic

Cilic

Djokovic

Pospisil

Murray

Federer

Federer

Figura 8: Resultados do torneio de Wimbledon 2015, a partir das quartas de final

É importante observar que não há necessariamente uma coincidência entre as listas dos jogadores mais cotados para atingirem as quartas de final e os maiores candidatos ao título. Isto se deve ao efeito do sorteio das chaves de cada torneio, que pode colocar adversários de diferentes tipos níveis de dificuldade no caminho dos favoritos, tornando a chave mais favorável para alguns deles.

Comparando com os resultados reais, para o torneio de Roland Garros, dos oito quadrifinalistas mais prováveis apontados pela metodologia, sete foram confirmados, enquanto que para Wimbledon, o número de coincidências foi de cinco. Quanto ao título, em ambas disputas os favoritos eram Novak Djokovic e Roger Federer, o que se confirmou na final de Wimbledon, vencida pelo sérvio então líder do *ranking*; em Roland Garros, contudo, Stan Wawrinka conseguiu superar as probabilidades adversas e vencer ambos os favoritos, sagrando-se campeão.

## 7 Conclusões

Este trabalho apresentou um estudo sobre a previsibilidade de vencedores em partidas de tênis, partindo de análises de históricos de desempenho dos jogadores, levando em consideração sua carreira, seu momento atual e sua *performance* nas diferentes superfícies. O problema da predição foi tratado por duas diferentes abordagens, a primeira por um sistema de inferência Fuzzy e a segunda por uma rede neural, ambas técnicas clássicas da inteligência computacional reconhecidas por sua eficiência e versatilidade em aplicações.

Os preditores apresentaram resultados satisfatórios, sempre superando o número de acertos obtido com a simples comparação de *rankings* e, em alguns casos, superando até mesmo o desempenho de apostadores. O aqui exposto, contudo, é parte de um modelo sujeito a muitas imperfeições, visto que é impossível quantificar dezenas de fatores de influência nos resultado de partidas, como o estado emocional momentâneo, as lesões, o apoio da torcida, o preparo físico, e a eventual falta de ritmo de jogo após ausência do circuito, adaptações a equipamentos, etc. No entanto, nota-se que há margens para melhoria nas predições, especialmente, ao se observar os resultados da rede neural para o conjunto de partidas disputadas em três *sets*, e quem um grande ganho foi conseguido.

Aprimoramentos serão buscados em trabalhos futuros, com o uso de informações de novas variáveis como os confrontos diretos entre jogadores e a premiação em dinheiro obtida na atual temporada ou em algum prazo determinado antecedendo o torneio sob análise, o que, de certa forma, representa com um peso maior as vitórias mais importantes, visto que os principais torneios oferecem premiações mais generosas e maiores valores monetários são concedidos em vitórias nas fases mais avançadas dos torneios. Essas informações, embora de obtenção mais difícil, podem permitir um melhor ajuste dos preditores. Outra sugestão é a de unificar

os resultados de diferentes sistemas em uma única previsão para cada partida, empregando, por exemplo, uma máquina de comitê ou outro tipo de agregador.

## Referências

- [1] ASSOCIATION OF TENNIS PROFESSIONALS. Official site of men's professional tennis. 2015. Disponível em: <a href="http://www.atpworldtour.com">http://www.atpworldtour.com</a>>. Acesso em: 1° out. 2015.
- [2] INTERNATIONAL TENNIS FEDERATION. 2015. Disponível em: <a href="http://www.itftennis.com">http://www.itftennis.com</a>. Acesso em: 15 set. 2015.
- [3] FORBES. *The world's highest-paid athletes*. 2015. Disponível em: <a href="http://www.forbes.com/athletes/list/">http://www.forbes.com/athletes/list/</a>>. Acesso em: 10 set. 2015.
- [4] GONZÁLEZ-DÍAZ, J.; GOSSNERB, O.; ROGERS, B. W. Performing best when it matters most: Evidence from professional tennis. *Journal of Economic Behavior & Organization*, Amsterdã, n. 84, p. 767–781, 2012. ISSN 0167-2681.
- [5] FERRAUTI, A. et al. Diagnostic of footwork characteristics and running speed demands in tennis on different ground surfaces. *Sport Orthopädie Traumatologie*, Frankfurt, n. 29, p. 172–179, 2013. Disponível em: <a href="http://dx.doi.org/10.1016/j.orthtr.2013.07.017">http://dx.doi.org/10.1016/j.orthtr.2013.07.017</a>>. Acesso em: 6 set. 2015.
- [6] KLAASEN, F.; MAGNUS, J. R. Forecasting the winner of a tennis match. *European Journal of Operational Research*, Amsterdã, n. 148, p. 257–267, 2003. ISSN 0377-2217.
- [7] MCHALE, I.; MORTON, A. A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, Amsterdã, n. 27, p. 619–630, 2011. ISSN 0169-2070.
- [8] CLOWES, S.; COHEN, G.; TOMLJANOVIC, L. Dynamic evaluation of conditional probabilities of winning a tennis match. In: AUSTRALIAN CONFERENCE ON MATHEMATICS AND COMPUTERS IN SPORT, 6. *Proceedings...* Gold Coast: 6M&CS, 2002. Disponível em: <a href="http://hdl.handle.net/10453/6673">http://hdl.handle.net/10453/6673</a>. Acesso em: 6 set. 2015.
- [9] KNOTTENBELT, W. J.; SPANIAS, D.; MADURSKA, A. M. A common-opponent stochastic model for predicting the outcome of professional tennis matches. *Computers and Mathematics with Applications*, Amsterdã, n. 64, p. 3.820–3.827, 2012. ISSN 0898-1221. Disponível em: <a href="http://dx.doi.org/10.1016/j.camwa.2012.03.005">http://dx.doi.org/10.1016/j.camwa.2012.03.005</a>>. Acesso em: 10 out. 2015.
- [10] CLARKE, S. R.; DYTE, D. Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research*, New Jersey, n. 7, p. 585–594, 2000. ISSN 1475-3995.
- [11] KLAASSEN, F.; MAGNUS, J. Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, n. 96, p. 500–509, 2001.
- [12] DEL CORRAL, J.; PRIETO-RODRIGUEZ, J. Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting*, Amsterdã, n. 26, p. 551–563, 2010. ISSN 0169-2070.
- [13] SCHEIBEHENNE, B.; BRODER, A. Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, Amsterdã, n. 23, p. 415–426, 2007. ISSN 0169-2070.
- [14] TENNIS DATA. *Tennis results and betting odds data*. 2015. Disponível em: <a href="http://www.tennis-data.co.uk/alldata.php">http://www.tennis-data.co.uk/alldata.php</a>>. Acesso em: 1° out. 2015.
- [15] HOLDER, R. L.; NEVILL, A. M. Modelling performance at international tennis and golf tournaments: is there a home advantage? *The Statistician*, New Jersey, n. 46, p. 551–559, 1997.
- [16] BARNETT, T.; POLLARD, G. How the tennis court surface affects player performance and injuries. *Medicine and Science in Tennis*, Renkum, v. 1, n.12, p. 34-37, 2007. ISSN 1567-2352.
- [17] WEISSTEIN, E. W. *Correlation Coefficient*. 2015. Disponível em: <a href="http://mathworld.wolfram.com/CorrelationCoefficient.html">http://mathworld.wolfram.com/CorrelationCoefficient.html</a>>. Acesso em: 20 out. 2015.

- [18] ZADEH, L. Fuzzy Sets. Information and Control, n. 8, p. 338-353, 1965. Disponível em <a href="http://dx.doi.org/10.1016/S0019-9958(65)90241-X">http://dx.doi.org/10.1016/S0019-9958(65)90241-X</a>. Acesso em: 15 set. 2015.
- [19] JANG, J.-S.; SUN, C.-T.; MIZUTANI, E. Neuro-Fuzzy and Soft Computing: a computational approach to learning and machine intelligence. Upper Saddle River: Prentice-Hall, 1997.
- [20] FERNANDES, M. A. Classificação de alvos utilizando atributos cinemáticos. Dissertação (Mestrado em Engenharia Aeronáutica e Mecânica) - Instituto Tecnológico de Aeronáutica, São José dos Campos, 2009.
- [21] SUGENO, M. et al. (Ed.). Industrial applications of Fuzzy control. New York: Elsevier Science, 1985.
- [22] BRAGA, A. P.; CARVALHO, A.; LUDERMIR, T. Redes neurais artificiais: teoria e aplicações. Rio de Janeiro: LTC, 2000.
- [23] CYBENKO, G. Approximation by superpositions of a sigmoidal function. Mathematics of Controls, Signals, and Systems, Berlim, Springer Verlag, n. 2, p. 303-314, 1989.
- [24] HAYKIN, S. Redes neurais: princípios e prática. Porto Alegre: Bookman, 2004.
- [25] LIMA, B. N. B. et al. Probabilidades no esporte. TRIM: revista de investigación multidisciplinar, Valladolid, n. 5, p. 39-53, 2012. Disponível em <a href="http://uvadoc.uva.es/handle/10324/11665">http://uvadoc.uva.es/handle/10324/11665</a>. Acesso em: 15 set. 2015.