



MÉTODO DE INVESTIGAÇÃO DA PERSONALIDADE DE ATLETAS OLÍMPICOS BRASILEIROS UTILIZANDO REDES NEURAIIS ARTIFICIAIS

Resumo - Vivenciamos um avanço exponencial tecnológico, sobretudo na aplicação da Inteligência Artificial, mais especificamente, técnicas de linguagem natural, contribuindo para o desenvolvimento de métodos pós-modernos de investigação da personalidade. Ao mesmo tempo, considera-se a história oral do atleta como uma fonte importante de construção do conhecimento a seu respeito, pois pode refletir experiências vividas ativamente por este sujeito, que transcende o aspecto objetivo, estando em constante processo de ressignificação, quer pelo imaginário popular, ou pela indústria cultural. Esta pesquisa trata-se de uma fase de desenvolvimento de uma metodologia de extração de informação e conhecimento, a partir da narrativa de atletas olímpicos brasileiros em entrevistas, submetidas a análises via Redes Neurais. A hipótese da pesquisa é de que seja possível identificar características associadas a traços de personalidade, por meio da aplicação de análises via Redes Neurais, para a extração automática de conhecimento e informação via associação de termos e frases descritores do modelo empírico dos cinco grandes fatores de personalidade presentes na literatura. Sugere-se que, assim que definido o método automático de extração de informação e conhecimento, investigar em um número muito maior de atletas, seja na mesma ou em distintas modalidades esportivas. Com este estudo, verificou-se relações interessantes entre sentenças que aparecem na narrativa dos atletas, em relação à traços de personalidade. Espera-se que este trabalho contribua para continuidade dos estudos com Redes Neurais, para investigação de traços de personalidade.

Palavras-chave: personalidade; avaliação; psicologia do esporte; redes neurais; inteligência artificial.

METHOD FOR INVESTIGATING THE PERSONALITY OF BRAZILIAN OLYMPIC ATHLETES USING ARTIFICIAL NEURAL NETWORKS

Abstract – We are experiencing an exponential technological advance, especially in the application of Artificial Intelligence, more specifically in natural language techniques, contributing to the development of post-modern methods for investigating personality. At the same time, the athlete's oral history is considered an important source of knowledge construction about him, as it can reflect experiences actively lived by this subject, which transcends the objective aspect, being in a constant process of ressignification, either by the imagination popular, or by the cultural industry. This research is a development phase of a methodology for extracting information and knowledge, from the narrative of Brazilian Olympic athletes in interviews, submitted to analysis via Neural Networks. The research hypothesis is that it is possible to identify characteristics associated with personality traits, through the application of analyzes via Neural Networks, for the automatic extraction of knowledge and information via association of descriptor terms and phrases of the empirical model of the five great factors of personality present in the literature. It is suggested that, once the automatic method of extracting information and knowledge is defined, investigate in a much larger number of athletes, either in the same or in different sports modalities. With this study, interesting relationships were verified between sentences that appear in the athletes' narratives, in relation to personality traits. It is expected that this work will contribute to the continuity of studies with Neural Networks, for the investigation of personality traits.

Keywords: personality; psychological assessment; sport psychology; neural networks; artificial intelligence.

MÉTODO PARA INVESTIGAR LA PERSONALIDAD DE LOS ATLETAS OLÍMPICOS BRASILEÑOS UTILIZANDO REDES NEURONALES ARTIFICIALES

Resumen - Estamos viviendo un avance tecnológico exponencial, especialmente en la aplicación de la Inteligencia Artificial, más concretamente en las técnicas del lenguaje natural, contribuyendo al desarrollo de métodos posmodernos de investigación de la personalidad. A su vez, la historia oral del deportista es considerada una importante fuente de construcción de conocimiento sobre él, ya que puede reflejar experiencias vividas activamente por este sujeto, que trasciende el aspecto objetivo, estando en un constante proceso de ressignificación, ya sea por la imaginación popular o por la industria cultural. Esta investigación es una fase de desarrollo de una metodología para la extracción de información y conocimiento, a partir de la narrativa de los atletas olímpicos brasileños en entrevistas, sometidas a análisis a través de Redes Neuronales. La hipótesis de investigación es que es posible identificar características asociadas a los rasgos de personalidad, mediante la aplicación de análisis vía Redes Neuronales, para la extracción automática de conocimiento e información vía asociación de términos descriptores y frases del modelo empírico de los cinco grandes factores de personalidad presente en la literatura. Se sugiere que, una vez definido el método automático de extracción de información y conocimiento, se investigue en un número mucho mayor de deportistas, ya sea en la misma o en diferentes modalidades deportivas. Con este estudio se verificaron interesantes relaciones entre las frases que aparecen en las narrativas de los deportistas, con relación a los rasgos de personalidad. Se espera que este trabajo contribuya a la continuidad de estudios con Redes Neuronales, para la investigación de rasgos de personalidad.

Palabras-clave: personalidad; evaluación psicológica; psicología del deporte; redes neuronales; inteligencia artificial.

Ivan Sant'Ana Rabelo

ivanrabelo@usp.br

Universidade de São
Paulo, Brasil

[http://dx.doi.org/
10.30937/2526-
6314.v5.id138](http://dx.doi.org/10.30937/2526-6314.v5.id138)

Recebido: 12 nov 2021

Aceito: 20 dez 2021

Publicado: 22 dez 2021

Introdução

O século XXI marca o avanço exponencial da tecnologia, sobretudo na utilização do que se denomina amplamente como Inteligência Artificial (IA), que permitiram investigações utilizando linguagem natural, *machine learning*, de diferentes tipos de dados e sobretudo metadados, apontam um novo marco na possibilidade de desenvolvimento de métodos pós-modernos de investigação da personalidade. Quando os dados disponíveis são textos escritos em língua natural, ou seja, dados não estruturados, o processo de extração de conhecimento é chamado de mineração de textos ou *text mining*. De maneira que a mineração de textos pode ser vista como uma especialização da mineração de dados, correspondendo à aplicação de um conjunto de técnicas para descoberta de conhecimento em textos¹.

O processo de mineração de textos pode ser visto como uma sequência de etapas genéricas, que devem ser instanciadas de acordo com os dados disponíveis e o conhecimento que se espera obter². As possíveis aplicações desse processo são diversas, como, por exemplo, análise exploratória de conteúdo textual, organização de coleções de documentos, análise de sentimentos e sistemas de recomendação³⁻⁵.

A aprendizagem de máquina é um campo da Inteligência Artificial que objetiva contribuir com o processo de aprendizagem do computador por meio de treinamento, e depois, o computador possa tomar decisões ou indicar respostas com base no que aprendeu, similar ao que ocorre com os humanos no processo de tomar decisões para realizar determinadas ações. Como ocorre na aprendizagem de matemática, na qual o aluno realiza muitos exercícios com o objetivo de aprender a resolver as questões matemáticas. Somente após a prática de muitos exercícios é que este aluno estará apto a fazer a prova e testar os seus conhecimentos aprendidos.

Assim, enquanto os humanos se apoiam em processos cognitivos para aprenderem a resolver seus problemas e realizar ações, máquinas precisam de algoritmos para que o um processo possa ocorrer. Os algoritmos de tomada de decisão nas máquinas dependem da análise de conhecimentos prévios, sendo fornecidos de algum tipo de base de dados ou sistema prévio para que possam ser feitas as inferências posteriores⁶.

Nas máquinas, dá-se o nome de processamento de linguagem natural (PLN) para essa subárea da Inteligência Artificial que atua na compreensão automática de linguagens humanas, adaptadas para serem manipuladas por computadores⁷. Entre as principais

aplicações desta tecnologia na atualidade estão os ‘*chatbots*’, que seriam os robôs de conversa, que são sistemas inteligentes projetados para conversarem com humanos em lojas on-line, operadoras de telefonia e televisão, e diferentes outras empresas nas quais esses softwares são empregados para atendimento ao cliente, sobretudo em triagens iniciais.

Portanto, o PLN diz respeito à interpretação de frases, tais como, ‘eu chorei de rir da piada’ e ‘a piada me fez chorar de raiva’. A primeira frase está mais relacionada a um afeto positivo (chorar teria a função de rir muito), enquanto a segunda frase pode se referir a uma emoção negativa. Na frase ‘A menina caiu, derrubou a boneca e quebrou a perna’ o sistema de PLN deverá ser capaz de identificar quem quebrou a perna, se foi a menina ou a boneca. Se até para nós humanos já pareça ser difícil obter essa resposta, para um computador identificar isso é algo muito complexo⁷.

Entre os problemas relacionados ao PLN está a questão da dimensionalidade do vocabulário. A comunicação humana é muito vaga às vezes, pois a maioria das pessoas acabam utilizando coloquialismos, abreviações e muitas vezes não há preocupação em corrigir erros ortográficos, principalmente no universo da internet e dos dados digitais⁶. Assim, o PLN visa promover um nível mais alto de compreensão da linguagem natural através do uso de recursos computacionais, com o emprego de técnicas para o rápido processamento de texto.

No aspecto da investigação de traços de personalidade, abordagens de traço e fatoriais de personalidade buscam uma pequena quantidade de dimensões que possam resumir os padrões consistentes de uma pessoa reagir. Allport⁸ postulou que os traços são aspectos invariantes que acompanham os aspectos mutáveis de uma pessoa. O modelo dos Cinco Grandes Fatores de Personalidade (CGF) é um exemplo de um fruto científico, dentro deste contexto, na busca de marcadores de traços de personalidade, dos últimos tempos. Tendo em vista que o modelo CGF tem suas origens na análise da linguagem utilizada para descrever pessoas, o uso de descritores de traços da linguagem natural tem sido defendido como a melhor estratégia para identificar fatores que permitam mensurar características de personalidade^{9,10}.

A origem do modelo CGF relaciona-se a um agrupamento contínuo de pesquisas sobre a personalidade humana, envolvendo especialmente as teorias fatoriais e também as de traço. O precursor de estudos do modelo foi McDougall, o primeiro estudioso a

apresentar uma explicação teórica da personalidade a partir de cinco fatores de personalidade, inspirando trabalhos de Thurstone, em 1934, no desenvolvimento de pesquisas para verificar empiricamente a adequação do modelo¹⁰.

Contudo, de acordo com Nunes, Hutz e Nunes¹¹ diferentes literaturas a respeito do modelo explicitam que foram demandados cinquenta anos para que os pesquisadores da personalidade dessem a devida atenção ao modelo CGF, e a partir disso vislumbrassem um proveitoso campo de trabalho para pesquisas. Goldberg¹² argumenta que se uma característica de personalidade for capaz de gerar diferenças individuais socialmente relevantes, as pessoas vão notar esta característica e, conseqüentemente, será inventada uma palavra ou expressão para descrever este traço¹³.

Em resumo, o modelo CGF segundo Nunes, Hutz e Nunes¹¹ pode ser descrito pelos fatores Extroversão, Socialização (Amabilidade), Realização (Conscienciosidade), Neuroticismo e Abertura à experiências. Vale ressaltar que a nomenclatura do CGF pode variar no Brasil, a depender do pesquisador. Neste parágrafo, optou-se por apresentar a nomenclatura proposta por Nunes et al.¹¹, mas foi mantido entre parênteses uma tradução um pouco mais literal, amplamente conhecida em pesquisas brasileiras de outros autores.

Portanto, abordagens de traço e fatoriais de personalidade tem buscado uma pequena quantidade de dimensões que possam resumir os padrões consistentes de uma pessoa reagir. Allport, em 1966, postulou que os traços são aspectos invariantes que acompanham os aspectos mutáveis de uma pessoa. O referido modelo, também conhecido na literatura internacional como *Five Factor Model* ou apenas por *Big Five*, tem sido vastamente estudado por ser uma referência teórica na descrição da personalidade de forma objetiva e clara^{14,15}. A aceitação do modelo também se relaciona ao vasto número de estudos e evidências de sua universalidade e aplicabilidade em contextos variados¹⁰.

Em tempo, vale destacar também a entrevista como uma fonte de informações a respeito das características pessoais e de personalidade. Considera-se a história oral do atleta como um fonte importante de construção do conhecimento a respeito dele e de sua personalidade, pois além de refletir experiências vividas ativamente por este sujeito, o atleta ocupa no mundo contemporâneo espaço que transcende o aspecto objetivo outrora atribuído, ou seja, é uma instituição em constante processo de ressignificação, quer pelo imaginário popular, ou pela indústria cultural¹⁶. Diferentemente de muitas áreas de relações e atuação do indivíduo, o esporte é um contexto que propicia, principalmente ao

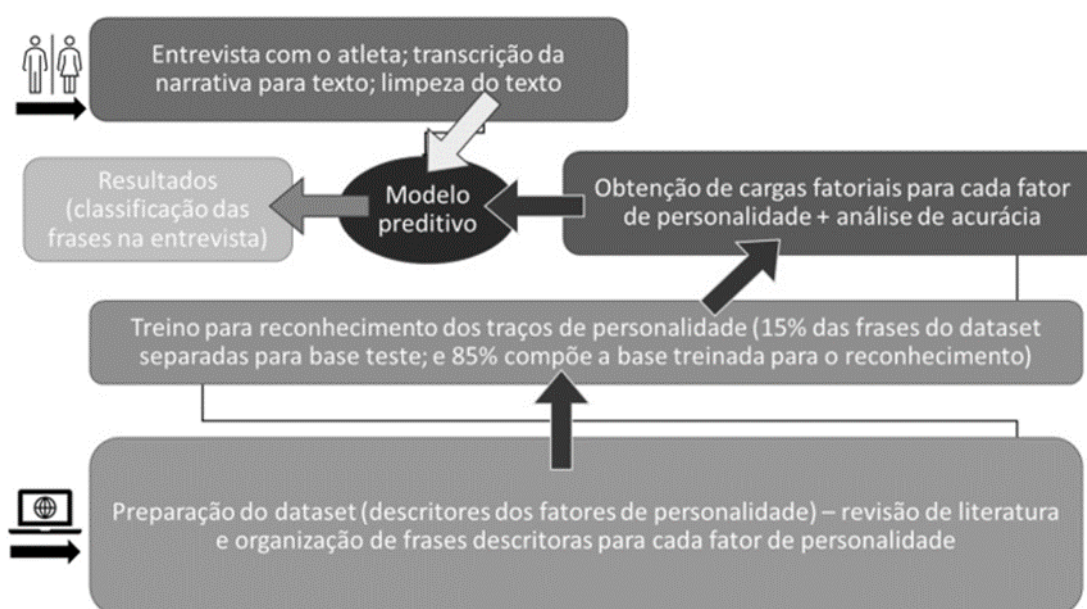
protagonista, as experiências afetivas e emocionais em graus elevados, que marcam primeiramente seu corpo e podem dar à memória um conteúdo ainda mais humano.

Com base no exposto, o objetivo deste artigo é apresentar um método de extração de conhecimento e informação, via associação de termos e frases descritores do modelo empírico dos cinco grandes fatores de personalidade presentes na literatura, para a identificação de traços nas narrativas da história de vida de atletas olímpicos brasileiros, submetidos à entrevista por narrativa oral gravada e depois transcritas em formato de texto.

Método

Trata-se de uma fase da pesquisa de desenvolvimento de uma metodologia de extração de informação e conhecimento associado a traços de personalidade de atletas. A hipótese da pesquisa é de que seja possível identificar características associadas aos traços de personalidade, a partir da aplicação de técnicas de mineração de textos em entrevistas, sendo desenvolvido e proposto o método apresentado na **Erro! Fonte de referência não encontrada.1**.

Figura 1. Proposta de método para extração de informação e conhecimento de traços de personalidade no modelo CGC de atletas por meio de IA.



Fonte: o autor.

A ideia central é realizar, por meio de aplicação de técnicas de *machine learning*, em português linguagem de máquina, mais especificamente processos de redes neurais artificiais, a extração de conhecimento e informação via associação de termos e frases descritoras do modelo empírico dos cinco grandes fatores de personalidade presentes na literatura, à partir da identificação nas narrativas da história de vida, obtidas de entrevistas transcritas.

Participantes

Participaram desta etapa da pesquisa de estudo exploratório de desenvolvimento do referido método de extração de informação e conhecimento 12 atletas olímpicos, da modalidade do tênis de mesa, de ambos os sexos, com idade média de 19,4 anos e desvio padrão de 4,7. Sugere-se que, após definido o método de extração de informação e conhecimento, investigar em um número muito maior de atletas, seja na mesma ou em distintas modalidades esportivas.

Procedimentos

A coleta das entrevistas foi realizada em um formato de entrevista aberta, com rebaixada intervenção do entrevistador, ou roteiro de perguntas previamente definido. A pergunta que inicia o diálogo é: ‘conte sobre sua história de vida’. Foi solicitado aos atletas participantes voluntários a assinatura do Termo de Consentimento Livre e Esclarecido. A pesquisa foi aprovada pelo Comitê de Ética em Pesquisa (CEP) sob número CAAE: 03499318.1.0000.5391 e número do parecer: 3.277.867.

Análise dos dados

Uma das ferramentas mais utilizadas para processamento de linguagem natural é o Natural Language Toolkit - NLTK (<http://www.nltk.org/>), que foi desenvolvido em Python e apresenta uma gama muito grande de recursos, tais como, classificação, tokenização, *stemming*, *tagging*, *parsing* e raciocínio semântico. Todas essas funções são utilizadas para análise dos textos. Outras vantagens do NLTK é a existência de versões para diferentes sistemas operacionais, seja Windows, Mac OS X e Linux, sendo também *open source* (códigos abertos, gratuitos), formando uma grande comunidade ativa de desenvolvedores que estão sempre atualizando e adicionando novos recursos.

O NLTK mostra-se útil para separar as sentenças em um parágrafo, separar as palavras dentro de cada sentença, reconhecer padrões no texto e criar modelos de classificação que permitam, por exemplo, identificar nomes próprios dentro de conjunto de dados. De acordo com Aranha¹⁷ o campo da Ciência da Computação e da Linguística abrange um conjunto de métodos formais para analisar textos e gerar frases em um idioma humano através do uso de programas computacionais. É possível ensinar um sistema a identificar regras e padrões estabelecidos pela linguagem na qual o texto está escrito, com duas ou três palavras de substantivos e substantivos próprios, não se retendo apenas a mineração de palavras separadamente.

Nesse método, são propostas as etapas descritas a seguir na tabela 1. Reforça-se que a mineração de dados é uma área multidisciplinar, que incorpora técnicas utilizadas em áreas como Inteligência Artificial, especialmente Aprendizado de Máquina, Base de Dados e Estatística. Com o processo de mineração de dados, ou seja, o processo de extração de conhecimento de grande volume de dados, busca-se a descoberta de conhecimento com a identificação de padrões novos, válidos, úteis e, muitas vezes, compreensíveis¹⁸.

Tabela 1. Protocolo de classificação – Método proposto

Etapa	Ação
Preparação do <i>Dataset</i>	<ul style="list-style-type: none">- Preparação do <i>Corpus</i> linguístico.- Para cada frase são identificados os <i>Pos-Tag</i>.- Remoção das <i>Stop Words</i>.- Remoção dos acentos.- Aplicado o <i>Stemmer</i> - converter para os radicais.- Montado o <i>DataFrame</i> (matriz) com a frase + <i>stemmer</i> + classe.
Base de treino e base de teste	<ul style="list-style-type: none">- Separação de base de treino e base de teste- Do <i>dataset</i> 15% das frases foram separadas para base teste.- Os demais 85% compõem a base treinada para o reconhecimento.
<i>Pipeline</i>	<ul style="list-style-type: none">- Conversão das frases em uma matriz de contagem de tokens.- Transformação da matriz em uma representação tf ou tf-idf normalizada (atribuição dos pesos para cada frase).- Calibração de probabilidade por regressão isotônica ou regressão logística.- Aplicado o <i>Linear Support Vector Machines</i> (LinearSVM) - estimativa da predição de carga fatorial em cada CGF.
Confiabilidade	<ul style="list-style-type: none">- <i>Cross-Validation</i> para obtenção da acurácia dos fatores.

Recordando, basicamente trata-se de um método que propõe a busca de frases presentes na narrativa da pessoa, coletada em entrevista, transcrita e submetida a comparação destas frases com uma base de informação (*dataset*) preparada previamente.

Esse *dataset* corresponde a um conjunto de frases que compõem um *corpus* com descritores dos cinco grandes fatores de personalidade. Utilizando procedimentos de aprendizagem de máquina, o sistema aprende a relacionar, frases descritoras em uma classificação de carga fatorial para cada um dos fatores do modelo CGF, ao mesmo tempo que, a cada entrevista submetida ao sistema, vai sendo ampliada sua acurácia.

Resultados

Inicialmente, ressalta-se que os indicadores de consistência, apresentaram calibração de 0,68, considerado interessante para os estudos preliminares com este método que está sendo desenvolvido. Na tabela 2 observa-se que as frases em sua maioria, carregam carga maior no fator Amabilidade, que condiz com o conteúdo e significado de cada uma das frase, de distintos atletas, assim como, apresentam alguns atletas com um número maior de frases que carregam mais fortemente neste fator, em relação a outros.

Também é possível verificar que a tabela 2 apresenta apenas os 15 primeiros itens com maior carga fatorial em Amabilidade de toda a base com as entrevistas de todos os atletas. Note que o nome dos atletas foi suprimido, substituídos pela letra A, seguido de um número que diferencia cada atleta participante em específico.

Tabela 2. Amabilidade (A)

Atleta ID	item linha	O	C	E	A	N
A11	os outros clubes o procuravam para aprender	0,02	0,01	0,11	0,87	0,01
A4	a gente vai te ajudar	0,04	0,07	0,03	0,84	0,02
A6	de repente com o intuito de não me ajudar(*)	0,04	0,07	0,03	0,84	0,02
A10	Para me ajudar e tal	0,04	0,07	0,03	0,84	0,02
A5	foi o clube que me formou como pessoa	0,04	0,03	0,07	0,81	0,04
A1	Hoje são outros tipos de seletiva(**)	0,04	0,05	0,09	0,80	0,02
A7	fora o futebol os outros esportes eram muito amadores(*)	0,01	0,02	0,13	0,73	0,11
A9	uma vida bem regrada(**)	0,01	0,08	0,03	0,72	0,17
A1	em relação a olimpíada(**)	0,09	0,06	0,03	0,72	0,10
A12	amigos de hoje são todos os meus amigos de infância	0,01	0,01	0,23	0,69	0,07

Qualitativamente, observam-se algumas confusões, que parecem estar refletindo um sentido inverso à descrição do fator Amabilidade, tal como nos itens com um asterisco na tabela (*), quais sejam, em A6 e A7. Assim como, algumas frases na tabela (**) que,

em princípio, parecem não ter uma relação perceptível diretamente com descrições na literatura a respeito do referido fator.

Este pode ser um ponto que precisará ser mais bem investigado no método, sendo que, entre as possibilidades de melhoria, pode estar a identificação de sentenças com cargas positivas e negativas, distintamente, para os fatores. Parece que o sistema ainda não está sendo capaz de distinguir tal aspecto, mas que, poderá ser conferido tal consistência, nos demais fatores que serão apresentados posteriormente, ainda neste manuscrito.

Tabela 3. Conscienciosidade (C)

Atleta ID	item linha	O	C	E	A	N
A9	trabalhar alinhado com a Confederação	0,01	0,93	0,03	0,01	0,01
A8	Mas eu vou continuar treinando e me dedicando a minha cidade	0,04	0,85	0,02	0,04	0,05
A1	tinha que trabalhar e tinha que estudar	0,11	0,83	0,02	0,02	0,02
A9	vamos treinar da melhor forma	0,04	0,82	0,03	0,05	0,05
A6	só que eu supri isso com a carga de treino e mais esforço e dedicação	0,02	0,81	0,06	0,01	0,10
A8	Minha mãe trabalhava muito(***)	0,03	0,81	0,04	0,08	0,04
A4	O objetivo deles era eu fazer algum exercício(***)	0,01	0,80	0,13	0,02	0,03
A5	Mas durante o ano eu fui melhorando os meus resultados	0,01	0,80	0,06	0,07	0,06
A7	está bem mais que profissional	0,01	0,80	0,03	0,11	0,05
A7	buscar fazer o melhor	0,04	0,80	0,10	0,05	0,02

Analisando também a interpretação semântica das sentenças da tabela 3 acima, com o referido fator Conscienciosidade, observa-se uma outra possível falha, que exigirão novos ajustes no método. Refere-se aos itens com três asteriscos (***) que indicam que o sistema ainda não se mostrou capaz de identificar quando o conteúdo semântico se referiu ao respondente (entrevistado) ou outros, como no caso ocorrido no A8, em que se tratava de traços da mãe e não do próprio sujeito.

Tabela 4. Extroversão (E)

Atleta ID	item linha	O	C	E	A	N
A6	eu falava eu quero estar lá	0,01	0,03	0,95	0,01	0,01
A12	eu tive uma infância muito alegre	0,02	0,03	0,90	0,02	0,02
A10	ele também quando eu falei para eles	0,03	0,04	0,90	0,01	0,02
A7	ao mesmo tempo ele falou que a gente tinha que ganhar(***)	0,01	0,04	0,86	0,01	0,07
A4	eu não sabia nem falar direito(*)	0,01	0,01	0,81	0,06	0,11
A12	eu tenho muitos primos	0,02	0,05	0,80	0,02	0,10

A10	eu lembro que o meu técnico falou assim	0,01	0,07	0,79	0,08	0,05
A5	Ele falou que era pra eu escolher(***)	0,01	0,09	0,79	0,01	0,11
A6	mas eu acho que ele falou isso na época para me cutucar	0,04	0,04	0,79	0,01	0,12
A3	e veio falar comigo(***)	0,01	0,01	0,78	0,01	0,19

Na tabela 4, aparece também a dificuldade em identificar quando o conteúdo semântico se referiu ao respondente (entrevistado) ou outros, isso pode ser observado, por exemplo, nos itens com três asteriscos. Assim como, no ID A4, parece que a sentença é em sentido oposto a carga no referido fator. Já na tabela 5 abaixo são apresentados os dados do fator Neuroticismo.

Tabela 5. Neuroticismo (N)

Atleta ID	item linha	O	C	E	A	N
A6	vou ficar triste é claro porque eu perdi o jogo	0,03	0,01	0,03	0,06	0,88
A10	quando você está jogando e você sente dor	0,01	0,01	0,07	0,04	0,87
A6	eu era ruim mesmo e só que aquilo surtiu o efeito contrário	0,01	0,05	0,05	0,03	0,86
A10	você acaba focando na dor	0,01	0,02	0,1	0,02	0,85
A6	aquilo não vai ficar me perseguindo eu tranquei a faculdade	0,03	0,01	0,07	0,04	0,85
A10	ou tantos jogos que você tem que ir com dor	0,02	0,01	0,07	0,04	0,85
A10	Você começa a parar para pensar na vida	0,01	0,06	0,07	0,02	0,85
A1	Eu infelizmente não peguei muito isso	0,01	0,01	0,03	0,10	0,85
A12	quando eu estava cansado	0,06	0,03	0,03	0,03	0,85
A2	você passa dificuldades assim no jogo	0,01	0,01	0,14	0,02	0,83

A próxima tabela (tabela 6) apresenta os resultados associados ao fator Abertura (O). Assim como já explicado nas tabelas anteriores, itens com sentido inverso à descrição do fator são destacados como itens com um asterisco na tabela (*), já (**) relacionam-se com itens que parecem não ter uma relação diretamente associada com descrições na literatura a respeito do referido fator. E os itens com três asteriscos (***) podem indicar que o sistema ainda não se mostrou ajustado para identificar quando o conteúdo semântico se referiu ao respondente (entrevistado) ou outras pessoas citadas pelo respondente na sua narrativa.

Tabela 6. Abertura (O)

Atleta ID	item linha	O	C	E	A	N
A8	E como eu era a mais nova(**)	0,96	0,01	0,01	0,01	0,02

A3	ele tinha mais uma mentalidade diferente(***)	0,94	0,01	0,01	0,02	0,02
A5	com uma cultura muito diferente do Brasil	0,90	0,02	0,02	0,03	0,03
A4	aceitar as coisas novas	0,90	0,03	0,02	0,04	0,01
A9	conhecer um novo idioma	0,88	0,01	0,05	0,01	0,06
A2	foi uma experiência nova	0,87	0,01	0,07	0,01	0,04
A2	foi uma experiência diferente	0,87	0,01	0,06	0,02	0,04
A9	a chance assim de conhecer novas culturas	0,86	0,01	0,03	0,02	0,07
A5	e eu fiquei curioso	0,85	0,03	0,03	0,03	0,07
A3	era um dos mais novos ali(**)	0,85	0,07	0,02	0,02	0,05

Estes são resultados preliminares no sentido de começar a organizar e, se necessário, desenvolver um *corpus* para contribuir na mineração de textos em entrevistas com os atletas, para identificação dos cinco grandes fatores de personalidade. De maneira que, a todo momento, sugere-se que sejam continuadas as pesquisas, assim como realizados novos estudos, com o objetivo de desenvolvimento de métodos de extração e organização de dados quantitativos, a partir de medidas qualitativas, tais como as entrevistas. Porém, conforme apresentado na tabela 7, ainda serão necessários ajustes no modelo, com o objetivo de ampliar sua sensibilidade à identificação das variáveis propostas, ainda que isso esteja previsto nas etapas de desenvolvimento da pesquisa, já que, por se tratar de um processo de aprendizagem de máquina, este *corpus* estará continuamente se qualificando.

Tabela 7. Sentenças que carregaram com múltiplos fatores

Atleta ID	item linha	O	C	E	A	N
A1	vamos entrar é pra ganhar	0,12	0,22	0,39	0,08	0,19
A3	a gente não tinha muita informação de técnica	0,01	0,02	0,50	0,07	0,40
A4	ele estava me ajudando muito	0,21	0,19	0,13	0,15	0,33
A6	foi muito difícil competir contra meu amigo	0,02	0,02	0,32	0,21	0,42
A8	me despertou o interesse em eu disputar uma olímpiada	0,29	0,05	0,39	0,16	0,11
A8	não que eu seja contra quem fuma	0,02	0,02	0,06	0,35	0,55
A9	pra enfrentar de igual pra igual os adversários	0,15	0,37	0,05	0,06	0,38

É esperado que distintas frases não estejam associadas fortemente a este ou aquele outro fator, exclusivamente. Afinal, isso mostra que o método é capaz de identificar em uma mesma sentença, diferentes traços presentes. O que é altamente adequado, haja vista, pode ser uma evidência de que o método, diferentemente de um teste de personalidade padronizado, em formato de escala, no qual foi previamente realizado uma análise fatorial

confirmatória, fixa que buscou definir cada frase de uma escala de personalidade como sendo deste ou daquele fator apenas. Estaria aqui uma vantagem de um método como este, utilizando *machine learning*, que é dinâmico e pode carregar cada sentença em diferentes fatores, simultaneamente.

Discussão

Considerando a subjetividade como tema diretamente relacionado à memória, diferentes interpretações das narrativas podem se fazer presente. Ao analisar as entrevistas de história de vida, o que se tem é não ter certeza sobre o fato relatado, se ele aconteceu do modo como foi narrado ou de outro jeito. Mas é exatamente a interpretação das histórias o elemento central das análises das entrevistas. Considerar quais experiências e fatos o atleta elege para ser narrado (uma vez que é sabido que essa história se tornará pública); o estilo de narrativa (expressivo, lacônico, com ou sem gestos etc.); o ponto de início da narrativa (a origem a partir dos pais, da vida esportiva, a infância); o local escolhido para a entrevista (casa, trabalho, rua, clube, espaços públicos); o encadeamento das ideias (o que é omitido ou silenciado deliberadamente); a organização do passado e como projeta o futuro; a intencionalidade de um sentido para a própria história de vida.

E nesse sentido, os atletas precisam ser investigados por meio de ferramentas construídas dentro do seu próprio contexto, pois, o uso de testes psicológicos de personalidade construídos fora do contexto esportivo, para populações gerais, quando aplicados em atletas apresentam resultados menos consistentes¹⁰. Isso pode estar relacionado ao fato de que o atleta está exposto a situações diferentes da população normativa do teste de personalidade.

Um exemplo refere-se a forma como é tratado ansiedade, a percepção da dor, o esforço etc., no contexto esportivo, pois muitas vezes os atletas estão expostos a situações que, na forma como é perguntado na escala, podem gerar respostas dúbias, entendimentos duplicados, sobre o mesmo fenômeno podendo o atleta responder com base em um aspecto ou em outro, diminuindo a confiabilidade dos resultados oriundos da escala, assim como, nestes mesmos estudos, podem ser observadas variações na estrutura fatorial da escala, em grupos específicos, como no caso de atletas¹⁰.

Em tempo, vale reforçar que, para avaliar e sobretudo interpretar resultados associados aos aspectos psicológicos no contexto do esporte, é imprescindível ter uma

ampla noção das questões relacionadas ao universo do esportista, desde noções de fisiologia, biomecânica, características específicas das modalidades, regras, questões socioculturais, a psicodinâmica de grupos esportivos entre outros aspectos¹⁹.

Estes são estudos preliminares no sentido de começar a organizar e, se necessário, desenvolver um corpus para contribuir na mineração de textos em entrevistas com os atletas, para identificação dos cinco grandes fatores de personalidade. Nesta pesquisa, parece que o método de separar uma base de treino e uma base de teste, com frases associadas aos fatores de personalidade no modelo dos cinco grandes fatores de personalidade, mostrou um promissor caminho, para a extração de informação e conhecimento de traços de personalidade, em entrevista, no formato de narrativa de história de vida. Isso se mostra interessante, para que os pesquisadores continuem a levantar informações teórico-empíricas sobre os fatores de personalidade, para alimentar o *dataset* que é utilizado para buscar as frases presentes na entrevista. Como conclusão deste estudo, foi possível observar uma acurácia de 68% na extração dos traços associados ao modelo CGF de investigação de personalidade, percentual considerado bastante aceitável para o objetivo proposto.

Porém, entre críticas existentes na ciência no que se refere à utilização de ferramentas de autorrelato para investigação de características associadas à personalidade, parece ficar claro haver distintos fatores externos que influenciam o comportamento das pessoas ao responderem um teste no formato de autopercepção, de julgamento sobre as próprias características. Seja, em primeiro lugar, pela possibilidade de baixo conhecimento sobre si mesmo e, portanto, o respondente poderia ter maior dificuldade de atuar julgar-se, mas ainda aspecto intervenientes diversos, tais como, a pressão de um superior pelo alto desempenho, a influência dos colegas no sentido da desejabilidade social, ou mesmo em razão de sistemas pessoais, mudanças tecnológicas, demandas da família, programas de treinamento e desenvolvimento, condições ambientais entre outras.

As variáveis intervenientes podem se apresentar das mais diversas fontes, tal como, o viés de resposta quando o sujeito avaliado está sofrendo alguma variação em razão do próprio ato de estar em processo de avaliação, de estar em testagem. Determinadas pessoas podem sofrer uma elevada influência da ansiedade, instabilidade emocional, quando se sentem pressionadas no processo avaliativo, como exemplo, alunos

que estudaram fortemente e assertivamente para o exame, mas que no momento da testagem, sentem-se em meio ao famoso “deu branco”, bloqueio de acesso à informação, recordação mnêmica. Sem contar as tentativas de manipulação da informação, muito presente em processo de recrutamento e seleção. A fome é um exemplo clássico de variável interveniente. Não se mensura diretamente a fome em um indivíduo ou em animal no laboratório, no entanto ela pode ser precisa e objetivamente relacionada com uma variável experimental como, por exemplo, o intervalo de tempo transcorrido desde a última vez em que o organismo recebeu comida.

Ainda, também dificuldades de compreensão leitora, como variável interveniente, ao responder escalas e medidas em formato de texto. São estes e outros desafios que elevam a importância da utilização de métodos e técnicas que possam utilizar menor pressão no processo investigação, de testagem.

E neste ponto, podemos destacar o resgate da entrevista, da história de vida, narrada pela pessoa, como uma possibilidade de compreender seu traços, atitudes, comportamentos e gostos, próprios, como mecanismo para investigação da personalidade. Valendo-se agora, da possibilidade do apoio de recursos tecnológicos, para auxiliar o profissional na coleta e análise de dados, diminuindo também uma outra variável interveniente conhecida, que é o viés ou falácia do investigador.

Sendo o avaliador um ser humano, está repleto de possibilidade de enganos diagnósticos, de percepção e interpretação que pode ser subjetiva e alterada em razão do momento, de fatores pouco controláveis. Ainda que seja prudente destacar que esta capacidade do ser humano em serem imprecisos, criativos e instáveis permitem que cada indivíduo seja alguém único. Objetiva-se auxiliar na acurácia do entendimento da características, ao mesmo tempo, com os avanços tecnológicos, possibilitar o cruzamento de grandes volumes de dados, em menor tempo, com mais acurácia, mesclando diferentes processamentos.

Conclusão

Com base no exposto, um aspecto conclusivo desta pesquisa refere-se à necessidade de ampliar os estudos realizados até o momento. Considera-se que parte da pretensão do projeto foi atendida, de propor o desenvolvimento de um método de análise de traços de personalidade nas narrativas dos atletas. Porém, a referida proposta de

método ainda tem caráter exploratório, haja vista, até o momento terem sido realizados estudos, aqui apresentados, de análises apenas em entrevistas de alguns conjuntos de atletas, mas ainda se sugere a continuidade destas análises, sua ampliação para mais atletas de mesma modalidade e para distintas modalidades, até alcançar a mineração de todas as entrevistas já transcritas dentro do projeto Memórias Olímpicas por Atletas Olímpicos Brasileiros²⁰.

Antes disso, é preciso aperfeiçoar o conjunto de dados (*dataset*) que compõe a base de treinamento, para que o *dataset* possa representar de forma mais acurada a maior parte da informação a respeito do fenômeno personalidade e seus subsistemas no modelo dos cinco grandes fatores. Assim como, da análise de características de personalidade nas suas nuances do contexto do esporte.

Também é preciso mais pesquisas para se ter clareza sobre quais são os melhores processos de Inteligência Artificial (IA) para serem aplicados para a etapa de *machine learning*. Sobretudo em razão das críticas à “caixa preta” em que os subsistemas de IA têm sido classificados, dada a dificuldade de “fiscalizar/acompanhar” o processamento, em razão de sua complexidade, por exemplo, nos sistemas chamados de “redes neurais”.

Mais especificamente em relação à utilização de métodos que aplicam inteligência artificial, tais como, aprendizagem de máquina, redes neurais e seus subsistemas, entre alguns pontos críticos, considerados por este trabalho como um “campo minado” do uso de algoritmos de IA, verificam-se diversificados dilemas éticos profundamente entrelaçados a este campo, por vezes ignorados, dentro da comunidade de tecnologia.

Afinal de contas, mesmo as melhores intenções não necessariamente são suficientes para garantir métodos e soluções livres de resultados indesejados ou não intencionais. Já que por se tratar, por exemplo, de um método que se propõe “ensinar” o computador, se este for “ensinado” por pessoas que pessoas podem codificar acidentalmente (ou não) preconceitos nos sistemas de IA, isso pode se tornar um grande problema. Sem contar que ainda outras pessoas má intencionadas também podem explorar as falhas dos modelos para objetivos que se quer ainda previstos.

De maneira que é importante não dar mais responsabilidade do que a máquina é capaz de suportar ou daquilo que nós humanos somos capazes de fiscalizá-la. Talvez não seja preciso temer uma superinteligência descontrolada que causa efeitos destrutivos, como vemos nos filmes futurísticos, mas considerar que tipo de habilidade deve ser dada

as máquinas. É preciso considerar que podemos decidir hoje um futuro irreversivelmente prejudicial, em que algoritmos adaptáveis e auto aperfeiçoados podem vir a guiar ou até ditar a maioria das decisões que definem a experiência humana.

As máquinas não substituirão os humanos completamente, é preciso enfrentar o ‘viés do algoritmo’. Segundo o artigo ‘*Garbage In, Garbage Out*’²¹, traduzido como: ‘entra lixo, sai lixo’, existem falhas inerentes das estruturas analíticas pelas quais entradas de baixa qualidade produzem invariavelmente resultados não confiáveis. O que isso quer dizer? Que o conjunto de dados que compõe a base de treinamento deve representar de forma adequada a maior parte da informação a respeito daquele fenômeno e seus subsistemas. Mas seria isso possível, sobretudo, no contextos das variáveis estudadas na psicologia, tal como os traços de personalidade?

Existem evidências científicas mostrando as desvantagens desse tipo de análise utilizando procedimentos de IA. Joy Buolamwini do MIT publicou uma pesquisa²² abordando a presença de preconceitos raciais e de gênero nos sistemas de IA fornecidos por algumas das maiores empresas de tecnologia do mundo, provavelmente, por causa dos dados estarem enviesados no conjunto de treinamento. É possível até imaginar os efeitos negativos de implantar esse modelo em um sistema de justiça criminal.

Contudo, isso não deve ser confundido com algoritmos adequadamente treinados, que não tratam necessariamente todos os grupos socioeconômicos, raciais, religiosos etc., da mesma maneira. Por exemplo, e se uma companhia de seguros de saúde identificasse uma enzima específica como sendo o preditor mais útil na identificação de pacientes que provavelmente precisam de uma cirurgia cara, e essa enzima é encontrada predominantemente em pessoas de uma determinada raça, seria justo que a pessoa dessa raça pague mais caro o seguro de saúde, mesmo que o valor de seguro tenha sido justificado pela presença da enzima e não pela cor de sua pele?²².

Algoritmos cada vez mais complexos, perdem seu processo de transparência, de acompanhamento humano, dada suas profundas camadas de análise, que geralmente geram resultados inesperados ou mesmo indesejados que não podem ser rastreados até sua origem. No entanto, não é como se a transparência representasse uma panaceia tecnológica. Suponha que hipoteticamente falando, se fosse provado estatisticamente que o gênero é o maior preditor de inadimplência do cartão de crédito, e todas as outras características tivessem relevância similar, essa disparidade no limite do crédito poderia

ser justificada ou considerada prudente? Não é possível pensar que faça sentido endossar qualquer tipo de discriminação baseada em gênero, ainda que tenham algoritmos que estejam propondo capacidade de predição a partir de dados coletados²².

Ainda é preciso ter certeza de que os objetivos perseguidos pela Inteligência Artificial estão de acordo com o bem-estar sustentável e a longevidade da humanidade. Embora seja verdade que os humanos também possam receber informações contraditórias (como ilusões de ótica, falsas memórias, mentiras etc.), é muito importante não dar, prematuramente, capacidades indevidas às máquinas.

Em conclusão, os dados certamente serão os recursos mais valiosos no mundo atual, e os sistemas de IA que os alavancam com mais eficiência serão o catalisador mais potente para a mudança que a humanidade já viu. Garantir que os desenvolvimentos sejam feitos com um olhar atento aos interesses de longo prazo da humanidade, deve ser uma das principais prioridades em todo o cenário tecnológico. A respeito do medo de que as máquinas substituam os profissionais e suas análises psicodinâmicas, ainda há muito o que evoluir, mas subestimar o apoio que as mais recentes ferramentas analíticas podem nos dar, e ao mesmo tempo passar prematuramente autoridade as máquinas, muito além da competência delas, seja por sua complexidade ou pelo seu treino inadequado, pode ser algo que precisa de maior e ampla discussão.

Referências

- 1 Aggarwal CC, Zhai C. *Mining Text Data*. New York (USA): Springer; 2012.
- 2 Rezende SO. *Sistemas inteligentes: fundamentos e aplicações*. São Paulo: Manole; 2003.
- 3 Sinoara RA, Camacho-Collados J, Rossi RG, Navigli R, Rezende SO. Knowledge-enhanced document embeddings for text classification. *Knowledge-Based Systems*. 2019;163:955-971. DOI: doi.org/10.1016/j.knosys.2018.10.026
- 4 Marcacini RM, Rossi RG, Matsuno IP, Rezende SO. Cross-domain aspect extraction for sentiment analysis: A transductive learning approach. *Decision Support Systems*. 2018;114:70-80. DOI: doi.org/10.1016/j.dss.2018.08.009
- 5 Sinoara RA., Antunes J, Rezende SO. Text mining and semantics: a systematic mapping study. *Journal of the Brazilian Computer Society*. 2017;23(1):1-20. DOI: doi.org/10.1186/s13173-017-0058-7
- 6 Rodríguez MM, Bezerra BLD. Processamento de linguagem natural para reconhecimento de entidades nomeadas em textos jurídicos de atos administrativos (portarias). *Revista de Engenharia e Pesquisa Aplicada*. 2020;5(1):67-77. DOI: doi.org/10.25286/repa.v5i1.1204
- 7 Bulegon H, Moro CMC. Mineração de texto e o processamento de linguagem natural em sumários de alta hospitalar. *Journal of Health Informatics*. 2010;2(2):51-56.

- 8 Allport GW. Traits revisited. *American Psychologist*. 1966;21:1–10.
- 9 Digman JM. Historical antecedents of the five-factor model. In: Costa PT, Widiger TA (Eds.), *Personality disorders and the five-factor model of personality*. Washington, DC: American Psychological Association. 2002. p. 17-22.
- 10 Rabelo IS. Investigação de traços de personalidade em atletas brasileiros: análise da adequação de uma ferramenta de avaliação psicológica [tese]. Escola de Educação Física e Esporte. São Paulo: Universidade de São Paulo; 2013.
- 11 Nunes CHSS, Hutz CS, Nunes MFO. Bateria fatorial de personalidade (BFP): manual técnico. São Paulo: Casa do Psicólogo; 2010.
- 12 Goldberg LR. The development of markers for the Big-Five factor structure. *Psychological Assessment*. 1992;4:26-42.
- 13 Hutz CS, Nunes CH, Silveira AD, Serra J, Anton M, Wieczorek LS. O desenvolvimento de marcadores para a avaliação da personalidade no modelo dos cinco grandes fatores. *Psicologia: Reflexão e Crítica* 1998;11(2):395-411.
- 14 McCrae RR, John OP. An introduction to the Five-Factor Model and its applications. *Journal of Personality*. 1992;60(2): 175-215.
- 15 Saucier G. Mini-Markers: A brief version of Goldberg’s unipolar Big-Five markers. *Journal of Personality Assessment*. 1994;63:506-516.
- 16 Rubio K. O Atleta e o mito do herói. São Paulo: Casa do Psicólogo; 2001.
- 17 Aranha CN. Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional [tese]. Pontifícia Rio de Janeiro: Universidade Católica do Rio de Janeiro – PUC; 2007.
- 18 Rubio K, Rabelo IS, Sinoara RA, Barbosa RS, Rezende SO. Identificação de personalidade de atletas olímpicos: uma análise exploratória de narrativa com mineração de textos. *Revista Brasileira de Psicologia do Esporte*. 2019;9(2):145-163.
- 19 Rabelo IS., Rubio K. Literatura científica sobre a mineração de textos aplicada à identificação da personalidade de atletas. *Olimpianos – Journal of Olympic Studies*. 2018;2(1):274-303. DOI:10.30937/2526-6314.v2n1.id37
- 20 Rubio K, Rabelo IS, Cruz RM. Avaliação de aspectos psicológicos em Educação Física e Esporte. In: Böhme MTS (org). *Avaliação e desempenho em Educação Física e Esporte*. São Paulo: Editora Manole; 2018.
- 21 Bininda-Emonds ORP, Jones KE, Price AS, Cardillo M, Grenyer R, Purvis A. Garbage in, Garbage out. In: Bininda-Emonds ORP (eds) *Phylogenetic Supertrees*. Computational Biology. New York (USA): Springer, Dordrecht; 2004.
- 22 Buolamwini J. How I'm fighting bias in algorithms [citado 18 fev. 2019]. 2017 Disponível em: https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms.