Avaliação da Personalidade (Big Five) a partir de posts no facebook usando o ColB5ERT

Ricardo Primi (Universidade São Francisco e EduLab21 Instituto Ayrton Senna)

Julho 2023

Abstract

Este artigo apresenta um estudo que utiliza posts do Facebook para avaliar a personalidade de indivíduos de acordo com o modelo Big Five (OCEAN: Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). Nós utilizamos o modelo ColB5BERT, uma adaptação do modelo de linguagem BERT, treinado com dados de posts do Facebook. Esta abordagem é uma tentativa de superar os limites dos métodos tradicionais de avaliação de personalidade que demandam tempo e estão propensos a distorções. Este trabalho discute a validade deste método e explora seus possíveis usos em várias áreas, como saúde mental, criação de conteúdo personalizado, pesquisa psicológica, recursos humanos e marketing

1 Introdução

O modelo Big Five é uma ferramenta consolidada na avaliação da personalidade, que busca descrever os traços e elementos fundamentais da personalidade (Soto & John, 2017). A avaliação da personalidade é realizada por meio de diversos métodos, como instrumentos de auto-relato, instrumentos de hetero-relato e entrevistas estruturadas.

As técnicas de avaliação baseadas em auto-relato são medidas indiretas, uma vez que exigem do avaliado o resgate de suas memórias de longo prazo para as experiências descritas nos itens, além da avaliação se o item descreve características de seu modo de agir. A resposta é então atribuída a um escore no item, geralmente numa escala Likert de 5 pontos (indicando de "discordo totalmente" a "concordo totalmente"). Veja Figura 1.

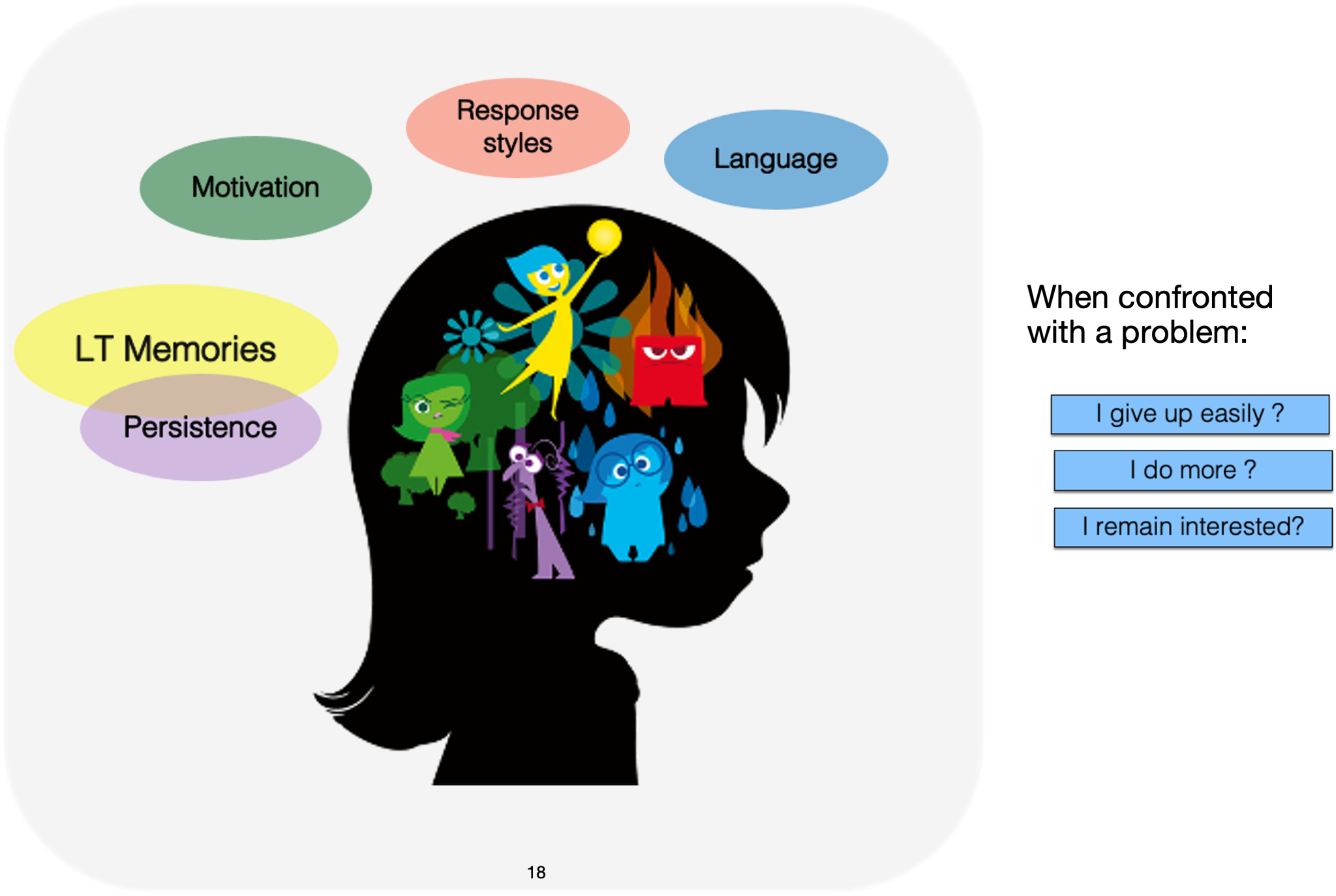


Figura 1 Ilustrando a medida indireta – mediada pelo sujeito – do traço persistência.

Essas medidas contrastam com medidas diretas, como os testes de inteligência, nos quais os avaliados devem resolver problemas que acionam a capacidade cognitiva a ser medida, demonstrando, diretamente e por meio de seu comportamento na situação de teste, uma amostra do traço que se quer medir, como a capacidade de raciocínio, por exemplo.

Na avaliação da personalidade, ainda não existe uma técnica viável de avaliação multidimensional do Big-Five baseada em observação direta. Pesquisadores e profissionais ainda confiam em instrumentos de auto-relato para esse fim. Esse tipo de avaliação demanda tempo e recursos do indivíduo e é propenso ao falseamento. Recentemente, pesquisadores têm explorado o uso de pegadas digitais, como os textos usados em redes sociais, junto com modelos de aprendizagem de máquina, construindo sistemas de avaliação dos fatores do Big Five (Basaran & Ejimogu, 2021; Neuman & Cohen, 2014; Stajner & Yenikent, 2020; Yamada et al., 2019).

O uso de textos para aferição da personalidade, investigado recentemente, depende da transformação de palavras, frases e parágrafos em vetores que representem essas unidades linguísticas. Tais indicadores podem ser produzidos a partir da contagem de termos e sua importância relativa em um texto (*bag of words*, tf, tf-idf) e/ou modelos baseados em redes neurais, como vetores estáticos densos de representação distribuída e, mais recentemente, vetores que capturam o uso de palavras dentro de um contexto utilizando modelos de linguagem (como o BERT). Os vetores são utilizados como variáveis independentes para a predição de medidas de personalidade, fazendo uso de modelos clássicos, como regressão, e de redes neurais LSTM, Convolucionais e Transformers.

Embora haja relativo sucesso na predição da personalidade utilizando textos, os resultados mostram um teto de correlação de 0,50 entre o sistema automático e os resultados concomitantes em testes de auto-relato consagrados em termos de precisão e validade. Adicionalmente, resta ainda saber se o controle de vieses de resposta dos instrumentos cujos escores são utilizados como variável dependente pode melhorar a validade dos modelos atuais. Por exemplo, será que o controle por aquiescência nos inventários pode fornecer um "*ground truth*" de maior qualidade, melhorando o dataset de treino e, consequentemente, produzindo modelos melhores de aferição de personalidade por textos?

Neste contexto, questiona-se se seria possível conhecer o perfil de características pessoais (personalidade, valores, fundamentos morais, interesses) a partir dos textos escritos pelas pessoas. Além disso, é relevante questionar se o uso de textos espontâneos, como posts e mensagens escritas nas mídias sociais, e textos eliciados, como autobiografias e respostas escritas para perguntas diretas sobre as características pessoais, podem ser usados de forma eficaz na avaliação da personalidade.

Nosso estudo explora a ideia de que os textos produzidos por uma pessoa podem ser uma fonte de observação direta mais válida para avaliar a personalidade. Um sistema automatizado como esse poderia ter uma amostra praticamente ilimitada do comportamento das pessoas sem necessidade de nenhum esforço por parte do avaliado.

Isso tem implicações importantes para várias áreas, como o monitoramento de saúde mental, a criação de conteúdo personalizado, a pesquisa psicológica, os recursos humanos e o marketing psicográfico. Este estudo procura investigar algumas questões amplas associadas a essa técnica: os textos produzidos espontaneamente serão boas amostras da personalidade de uma pessoa? Qual o tamanho desse texto para se ter uma visão compreensiva da personalidade? Quais tipos de textos serão mais informativos? Estes problemas guiarão o restante deste estudo.

2 O sistema colB5BERT = ColBERT + B5

Um sistema de busca é uma aplicação de software desenvolvida para localizar informações armazenadas em um banco de dados que sejam relevantes para uma consulta, ou 'query', do usuário. Essas informações podem incluir documentos de texto, imagens, entre outros tipos de dados. A eficácia de um sistema de busca é determinada pela precisão, relevância e utilidade dos resultados retornados pelos seus algoritmos.

Sistemas de busca podem operar em diversos níveis de sofisticação. Alguns utilizam correspondência exata de termos, como o BM25, um algoritmo baseado em modelo probabilístico que considera a frequência dos termos nos documentos e em toda a coleção de documentos. Este algoritmo retorna documentos que contêm as palavras da consulta exatamente como inseridas pelo usuário. Embora seja eficiente para recuperar documentos que contêm a consulta exata, este método apresenta limitações: ignora a semântica dos termos da consulta e não é capaz de lidar com sinônimos ou termos relacionados. Além disso, a correspondência exata não considera a relevância do contexto em que a consulta aparece no documento.

O ColBERT (*Contextualized Late interaction over BERT*) é um modelo baseado em redes neurais que utiliza a arquitetura BERT (*Bidirectional Encoder Representations from Transformers*). Esse modelo consegue captar a semântica dos termos, identificando palavras diferentes que são sinônimas. Mais importante ainda, o modelo leva em conta o significado dos termos no contexto em que aparecem, criando vetores densos contextuais para as palavras.

O ColBERT processa a consulta e o documento (ou passagens) separadamente pelo BERT, gerando assim um conjunto de vetores para cada termo da consulta e cada palavra do documento. Em seguida, ele cria uma matriz de semelhanças (medidas pelo cosseno dos vetores) entre a combinação de cada palavra da consulta com cada palavra do documento, a chamada "late interaction". A partir dessa matriz, o modelo agrega a informação, primeiramente pelo maior coeficiente em cada linha, e em seguida calculando a média desses números para produzir um escore de relevância do documento para a consulta.

O ColBERT resolve limitações significativas dos métodos de busca tradicionais. Ele é capaz de entender a semântica e o contexto dos termos, melhorando a eficácia na recuperação de documentos relevantes para a consulta. Além disso, ele lida com a ambiguidade linguística, considerando o contexto em que os termos aparecem.

Mas como isso se relaciona com a avaliação de personalidade?

Tradicionalmente, a avaliação de personalidade é realizada por meio de auto-relatos, solicitando que os indivíduos respondam a uma série de perguntas ou afirmações. Neste contexto, essas afirmações podem ser consideradas "queries" que visam capturar diferentes aspectos da personalidade. As respostas dos indivíduos em uma escala Likert representam um escore de quão relevante é cada afirmação para descrevê-los. Esses escores são então agregados para fornecer uma medida dos traços de personalidade do indivíduo.

Essa abordagem exige que o indivíduo acesse sua "memória de longo prazo", ou seja, a representação acumulada de suas experiências, comportamentos e sentimentos ao longo da vida, para responder às queries de forma apropriada. Assim, podemos ver essas representações acumuladas de experiências como documentos que descrevem a vida do indivíduo.

Ao entender o indivíduo como uma "coleção de documentos", oferecemos uma nova maneira de abordar este problema por meio de sistemas de busca. Cada texto, post, tweet ou outra produção linguística pode ser visto como um "documento" individual, uma observação direta dos comportamentos do indivíduo.

A hipótese central deste estudo é que as passagens textuais produzidas por um indivíduo oferecem uma representação mais rica e autêntica de sua personalidade do que os auto-relatos. Essas passagens podem ser analisadas para produzir uma medida de personalidade que evita as limitações cognitivas impostas pelos métodos tradicionais de auto-relato.

Vendo os itens de um teste de personalidade como queries e os textos de um indivíduo no Facebook, por exemplo, como documentos permite-nos abordar este tema como um problema de busca, onde podemos usar o ColBERT para calcular um escore de relevância de cada item em relação aos posts. No entanto, no contexto da avaliação da personalidade, estamos mais interessados nos escores de relevância de cada item ou query, pois esses podem ser agregados para fornecer uma medida dos traços de personalidade, semelhante aos testes de personalidade convencionais.

Neste estudo, propomos o uso do modelo ColBERT para extrair representações vetoriais dessas passagens de texto. A "relevância" de uma passagem para uma determinada query é calculada como a semelhança entre a representação vetorial da passagem e a representação vetorial da query. A medida da personalidade é então calculada como a agregação desses escores de relevância. Daí surgiu o nome do modelo: ColB5BERT, uma combinação de ColBERT e B5 (Big Five).

Em resumo, a ideia central deste trabalho é utilizar o modelo de linguagem ColBERT para transformar o texto produzido por um indivíduo e os itens de personalidade em uma série de passagens representativas e queries respectivamente. Isso evita a necessidade de auto-relato e os desafios associados à lembrança e ao processamento da informação pessoal. Acreditamos que essa abordagem oferece uma maneira nova e promissora de avaliar a personalidade, com potencial para superar os métodos tradicionais.

3 Método

Para realizar esta pesquisa, utilizamos uma base de dados composta por posts do Facebook e itens de três testes de personalidade em português. A base de posts foi dividida em chunks de 250 tokens, resultando em um total de 11537 posts (Santos & Paraboni, I2019). A base de itens consiste em 415 itens de 3 testes de personalidade em português avaliando o big-five.

O modelo utilizado para a transformação de texto em vetores numéricos foi o BERTimbau, uma versão do BERT pré-treinada em português (Souza, Nogueira, & Lotufo, 2020).

Inicialmente separamos os indivíduos com escore acima do percentil 75 e abaixo do percentil 25 como pessoas típicas com traços salientes de personalidade tanto no polo positivo (por exemplo, extroversão) como baixo (como por exemplo introversão ver Figura 2). Isso foi possível pois na base de dados estava disponível o escore dos sujeitos no teste BFI que mede o big five via auto relato.

Gráfico, Histograma

Descrição gerada automaticamente

**Figura 2.** Organização dos exemplos positivos para treinamento do ColB5BERT

Os posts desses indivíduos foram combinados com os itens correspondentes aos fatores que tinham escore alto consistindo em pares positivos item-post ou *query-document*. A combinação para exemplos negativos foi feita pela técnica conhecida como “*in batch negatives*”, isto é, para cada exemplo do lote, os outros posts correspondentes há outros itens de outros traços de personalidade serviram como exemplos negativos item-post ou query-document.

Com esse esquema iniciou-se os parâmetros do BERTimbau e executou-se o ajuste fino dos parâmetros do modelo (fine tunning).

A função de perda estava baseada na similaridade entre um vetor único correspondente ao item e um vetor único correspondente ao post. Esses vetores correspondiam a média dos vetores de cada token que compunham o item e o post (ver Figura 3). O modelo foi otimizado por meio desse treinamento para produzir similaridade máxima para os exemplos positivos, isto é, posts de pessoas extrovertidas com itens de extroversão, por exemplo, e similaridade baixa caso sejam exemplos de uma pessoa que não tem esse traço elevado.

O treinamento foi feito no google colab pro+ com uma GPU A100 com 30GB de memória e CPU de 80Gb de RAM.

Na fase de inferência com a base de teste, cada combinação post-item recebeu um escore de similaridade permitindo-se calcular os escores por fator pela média dos escores por cluster de itens de cada fator do Big 5. Esses escores foram considerados no nível das interações post – fator e por sujeito.

Interface gráfica do usuário

Descrição gerada automaticamente

**Figura 3.** Similaridade para a *loss fiunction*: o modelo foi treinado para maximizar a similaridade nos exemplos positivos

3 Data set

O dataset utilizado para treinar o ColB5BERT consiste em 415 itens de 3 testes de personalidade em português. Além disso, o dataset inclui posts do B5 corpus, que inclui dados de 1,019 pessoas e 194,382 posts, totalizando 2,219,585 tokens. Os posts foram divididos em chunks de 250 tokens, resultando em um total de 11,537 posts (Santos & Paraboni, I2019).

Depois de filtrados os casos de *match* positivo a base de treino ficou composta de 993.759 interações item-post de N=868 pessoas. E a base de teste 158.594 interações de N=152 pessoas.

Esses exemplos foram ordenados aleatoriamente para garantir em batches de 32 casos de interação item-post tivéssemos pessoas diferentes e itens diferentes consequentemente tendo uma amostra adequada de exemplos negativos.

Esse dataset contém escores de cada indivíduo no teste BFI permitindo que calculássemos a correlação entre o escore automático e o escore do indivíduo em um teste de autorrelato como métrica para avaliar a qualidade do modelo. Quanto mais essa métrica se aproximar de 1 (entre 0,70 a 0,90) melhora será o modelo já que atingiu um benchmark de precisão adequada de teste-reteste.

4 Experimentos

Neste estudo, realizamos experimento preliminar para verificar se é possível identificar traços de personalidade a partir de textos produzidos pelas pessoas. Para treinar o modelo ColB5ERT, utilizamos a técnica de ajuste fino (finetunning), que consiste em ajustar um modelo pré-treinado a uma nova tarefa específica.

Para treinar o ColB5BERT, como explicado, foram criados exemplos relevantes e não relevantes, ordenando os dados de forma aleatória para garantir que haveria sujeitos diferentes no mesmo lote. Assim os negativos foram obtidos das outras postagens no mesmo lote de treinamento, enquanto os exemplos positivos consistiam dos posts dos usuários aos itens do teste de personalidade condizentes com seu resultado no BFI. Assim durante o processo de treinamento tínhamos exemplos negativos (exemplos que não correspondem ao traço de personalidade que estamos tentando prever) e exemplos positivos (exemplos que correspondem ao traço de personalidade).

O modelo foi treinado por 3 épocas, levando cerca de 6h40m por época (cada época representando uma passagem completa pelos dados de treinamento). A Tabela 1 apresenta os resultados da função de perda. Nota-se que l*oss* na base de teste foi aumentando a cada época indicando que o modelo já estava apresentando *overfiting* já na primeira época.

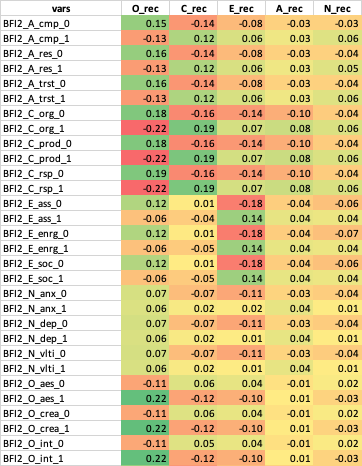
**Tabela 1.** Resultados da loss para cada época

|  |  |  |
| --- | --- | --- |
| Época | Loss Train | Loss teste |
| 1 | 3.4 | 4.2 |
| 2 | 2.3 | 6.5 |
| 3 | 1.8 | 6.2 |

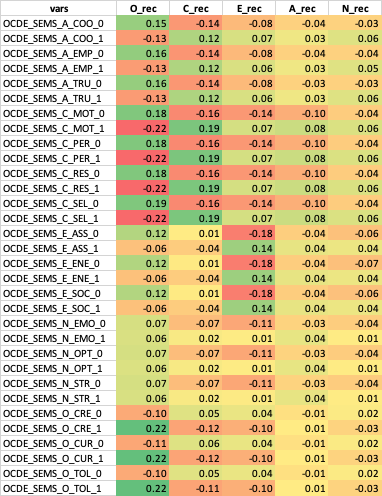
Na fase de inferência, utilizamos o state\_dict do modelo ao final da primeira época. Em seguida, cada combinação post-item na base de teste foi pontuada e agregada por sujeito X posts X item, resultando em uma média por cluster de itens de cada fator do B5. Assim avaliamos em que medida os escores automáticos estavam correlacionados com os escores do BFI considerando somente a base de teste.

As correlações estão apresentadas nas Tabelas 2-7.

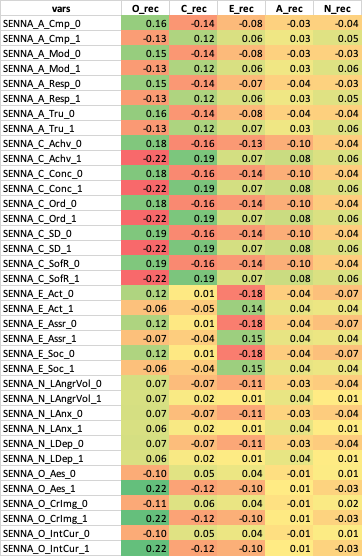
**Tabela 2.** Correlações das escalas do BFI2 calculados com o ColB5BERTcom os escores do BFI controlados por acquiescência no nível das respostas



**Tabela 3.** Correlações das escalas do SEMS calculados com o ColB5BERTcom os escores do BFI controlados por acquiescência no nível das respostas



**Tabela 4.** Correlações das escalas do SENNA calculados com o ColB5BERTcom os escores do BFI controlados por acquiescência no nível das respostas



**Tabela 5.** Correlações das escalas do BFI2 calculados com o ColB5BERTcom os escores do BFI controlados por acquiescência no nível dos sujeitos

Tabela

Descrição gerada automaticamente com confiança média

**Tabela 6.** Correlações das escalas do SEMS calculados com o ColB5BERTcom os escores do BFI controlados por acquiescência no nível dos sujeitos

Tabela

Descrição gerada automaticamente

**Tabela 7.** Correlações das escalas do SENNA calculados com o ColB5BERTcom os escores do BFI controlados por acquiescência no nível dos sujeitos

Tabela

Descrição gerada automaticamente

Por fim fizemos a regressão predizendo o escore global do Big Five a partir dos escores de todos os três instrumentos. Os resultados são apresentados na Tabela 8

**Tabela 8.** Regressão múltipla predizendo os escores do BFI controlados por acquiescência no nível dos sujeitos a partir dos escores automáticos calculados com o ColB5BERT

Tabela

Descrição gerada automaticamente

5 Conclusão

Este estudo apresentou uma nova abordagem para a avaliação da personalidade utilizando posts do Facebook. Os resultados preliminares são promissores, em alguns casos atingindo a métrica máxima observada na literatura, sugerindo que a abordagem proposta pode ser uma alternativa válida para os métodos tradicionais de avaliação da personalidade.

Esse estudo mostrou que a utilização de modelos de linguagem como o BERT, ajustados para tarefas específicas como a identificação de traços de personalidade, pode ser um caminho promissor para se criar um sistema de avaliação da personalidade a partir de textos produzidos por uma pessoa.

No entanto, ainda existem várias questões que precisam ser respondidas. Por exemplo, ainda não está claro quais tipos de textos são mais informativos da personalidade. Quais traços são mais validadente capturados e qual o tamanho ideal do texto necessário para obter uma visão abrangente da personalidade de uma pessoa.

6 Trabalhos Futuros

Em trabalhos futuros, planejamos investigar mais detalhadamente as questões levantadas neste estudo. Em particular, queremos entender melhor quais posts são mais informativos para quais dos fatores do big five. Mais uma direção seria usar o método em datasets maiores e com exemplos negativos de maior qualidade.

Outra direção potencial para a pesquisa futura é a investigação da ética do uso de dados de mídia social para a previsão da personalidade. Embora essa técnica possa ter aplicações úteis, como o monitoramento da saúde mental e a criação de conteúdo personalizado, também existem preocupações significativas em relação à privacidade e ao consentimento dos usuários.

Referencias

Basaran, S.; Ejimogu. O. H (2021) A Neural Network Approach for Predicting Personality From Facebook Data. Sage Open 1(15). https://doi.org/10.1177/21582440211032156

Borkenau, P; Mosch, A; Tandler, N; Wolf, A. (2014) Accuracy of Judgments of Personality Based on Textual Information on Major Life Domains. Journal of Personality 84(2) DOI: 10.1111/jopy.12153

Hernández-Dorado, A; Vigil-Colet, A; Lorenzo-Seva, U & Ferrando, P. J. (2021) Is Correction for Aquiescence Increasing the External Validty of Personality Test Scores? Psicothema Vol 33, n°4 639-646 doi: 10.7334/psicothema2021.131

Hommel, B. E., Wollang, F.-J. M., Kotova, V., Zacher, H., & Schmukle, S. C. (2022). Transformer-based deep neural language modeling for construct-specific automatic item generation. Psychometrika, 87, 749–772. https://doi.org/10.1007/s11336-021-09823-9

Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2020). Recent Trends in Deep Learning Based Personality Detection. Artificial Intelligence Review, 53(4), 2313–2339. https://doi.org/10.1007/s10462-019-09770-z

Neuman, Y.; Cohen, Y (2014) A Vectorial Semantics Approach to Personality Assessment. Scientific Reports DOI: 10.1038/srep04761

Plank, B., & Hovy, D. (2015). Personality Traits on Twitter—Or—How to Get 1,500 Personality Tests in a Week. Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 92–98. https://doi.org/10.18653/v1/W15-2913

Santos, V. G.; Paraboni, I. (2022). Myers-Briggs personality classification from social media text using pre-trained language models. Journal of Universal Computer Science, 28(4).

Santos, W. R. dos, & Paraboni, I. (2019). Personality facets recognition from text (arXiv:1810.02980). arXiv. https://doi.org/10.48550/arXiv.1810.02980

Siddique, F. B., Bertero, D.; Fung, P. (2019) GlobalTrait: Personality Alignment of Multilingual Word Embeddings. Proceedings of the aaai conference on artificial intelligence (Vol. 33, No. 01, pp. 7015-7022). https://doi.org/10.48550/arXiv.1811.00240

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. Journal of Personality and Social Psychology, 113(1), 117–143. https://doi.org/10.1037/pspp0000096

Souza, F. Nogueira, R. & Lotufo, R. (2020) BERTimbau: pre- trained BERT models for Brazilian Portuguese. In 9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear).

Stajner, S., & Yenikent, S. (2020). A Survey of Automatic Personality Detection from Texts. Proceedings of the 28th International Conference on Computational Linguistics, 6284–6295. <https://doi.org/10.18653/v1/2020.coling-main.553>

Trull, T. J.; Widiger, T. A., Useda, J. D.; Halcomb, J; Doan, B.; Axelrod, S. T. (1998) A Structured Interview for the Assessment of the Five-Factor Model of Personality. Psychological Assessment 10(3). DOI: 1040-3590/98/J3.00

Xie, D. Cobb, C. L (2020) Revised NEO Personality Inventory. The Wiley Encyclopedia of Personality and Individual Differences: Measurement and Assessment. 335-350. DOI:10.1002/9781119547167

Yamada, K. Sasano, R. Takeda, K (02019) Incorporating Textual Information on User Behavior for Personality Prediction. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 177–182 Florence, Italy.