

Modelo computacional para identificação de perfil de personalidade baseado em textos educacionais

Jose Antonio Buiar

Universidade Tecnológica Federal do Paraná
Curitiba, Paraná, Brasil
buiar@utfpr.edu.br

Luis Oliveira

Universidade Federal do Paraná
Curitiba, Paraná, Brasil
lesoliveira@inf.ufpr.br

Andrey Ricardo Pimentel

Universidade Federal do Paraná
Curitiba, Paraná, Brasil
andrey@inf.ufpr.br

Zenaide Carvalho da Silva

Universidade Federal do Sul e Sudeste do Pará
Marabá, Pará, Brasil
zenaide.silva@unifesspa.edu.br

ABSTRACT

One of the challenges found in education is the need of student's particularities adaptation, including affective aspects and personality. This article presents the Parameterizable Personality's Profile Identification Model (IP3), that uses the information found in texts, elaborated by students, to proceed the automated inference of personality traits, in accordance to the Big Five model. The findings of this research show this model's viability applied in learning environments.

Keywords

personality recognition; big five personality model; natural language processing; learning environment.

RESUMO

Um dos desafios encontrados da educação é a necessidade de adaptação às particularidades dos alunos, considerando sua personalidade e os aspectos afetivos envolvidos. Este trabalho apresenta o Modelo de Identificação de Perfil de Personalidade Parametrizável (IP3), que utiliza as informações presentes nos textos elaborados pelos alunos, para a realização da inferência automatizada dos traços de personalidade, de acordo com o modelo Big Five. Os resultados apresentados no experimento realizado demonstraram a viabilidade da utilização deste modelo em ambientes educacionais.

Palavras-chave

identificação de personalidade; modelo big five de personalidade; processamento de linguagem natural; ambiente de aprendizagem.

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Times New Roman 8-point font. Please do not change or modify the size of this text box.

Each submission will be assigned a DOI string to be included here.

ACM Classification Keywords

Applied computing~Interactive learning environments,
Applied computing~Learning management systems.

INTRODUÇÃO

O processo de ensino aprendizagem, nas últimas décadas, tem recebido uma crescente contribuição por parte das investigações realizadas na área da psicologia, as quais vêm propondo uma mudança significativa para as práticas escolares, visto que essas reflexões têm provocado um deslocamento no eixo-pedagógico, mudando a valorização de como e quem ensina, para a preocupação de quem aprende e de como se aprende. Neste processo mais participativo, começam a ser relevantes as diferenças existentes entre os indivíduos, do ponto de vista do processo educacional. Conforme verificado por [24], considerar as diferenças é encontrar situações de aprendizagem ótimas para cada aluno, buscando uma educação individualizada.

Uma das formas de verificação da diferença entre os alunos, é a identificação do perfil de personalidade. Esta identificação pode ajudar o professor como por exemplo: na formação de grupos colaborativos [8]; na identificação do estilo de aprendizagem para utilização de interfaces adaptativas [Omitido 2017b]; na verificação da correlação com satisfação acadêmica [4], na recomendação de estratégias pedagógicas [17] e na identificação de dominantes e seguidores em grupos de colaboração [29].

O método manual mais utilizado para identificação da personalidade é a utilização de questionários de avaliação de personalidade, também conhecidos como inventários [18]. A utilização dos questionários de avaliação para alunos, na forma de questionários de autoavaliação ou aplicados por especialistas, também apresenta os desafios práticos de utilização por ser um método intrusivo. Nem sempre existem profissionais habilitados a aplicar estes questionários nos alunos, bem como não se pode assegurar que os alunos expressem naturalmente suas respostas em formulários *online*, aplicados à distância.

Outra forma de investigar o perfil de personalidade dos alunos, é buscar a identificação dos traços de personalidade por meio das pistas encontradas em suas manifestações

naturais, em textos redigidos por estes durante as atividades realizadas no decorrer do curso. A utilização das pistas encontradas no texto para a realização da identificação da personalidade automatizada, não especificamente no ambiente educacional, mas de forma mais generalizada, foi comprovada inicialmente pelos estudos de [2,14,20].

Este trabalho apresenta um modelo de identificação da personalidade dos alunos, utilizando aprendizado de máquina a partir de um conjunto de técnicas de representação e estruturação do texto. Estas técnicas englobam utilização do léxico LIWC e os modelos de *Word Embedding* e *Word2Vec*. O modelo apresentado possibilita a escolha de bases diversas de treinamento, bem como de classificadores para a realização da inferência do perfil de personalidade dos alunos, de acordo com o modelo *Big Five* de personalidade, a partir dos textos obtidos nas atividades educacionais. Este modelo foi aplicado em um experimento realizado em um ambiente educacional sendo os resultados obtidos apresentados.

TRABALHOS RELACIONADOS

Apesar da verificação de estudos recentes que tratam da identificação do perfil de personalidade a partir do texto, são poucos os trabalhos que abordam esta identificação com objetivos educacionais no Brasil. Podem ser destacadas as iniciativas de [26] que investiga a identificação da personalidade a partir do teclado, [8] que investiga a formação de grupos de colaboração entre alunos baseado na identificação do perfil de personalidade e a pesquisa conduzida por [Omitido 2017a] que aborda a identificação do estilo de aprendizado a partir do perfil de personalidade reconhecido no texto dos alunos.

Dentre os trabalhos internacionais de identificação a partir de textos, [10] utilizou técnicas de *n-gram* para a identificação da personalidade a partir do texto, [28] procurou investigar a utilização de técnicas de redução de características nos processos de identificação e [15] utilizou redes neurais convolucionais. Estes autores utilizaram a base ESSAYS [23] para a realização destes experimentos.

PERSONALIDADE

Uma das principais abordagens utilizadas para o estudo da personalidade humana é a teoria dos traços de personalidade. Esta teoria está baseada na mensuração de traços que podem ser definidos como padrões habituais de comportamento, pensamento e emoção [13]. O conceito de traços de personalidade é a modelagem que mais efetivamente prediz aspectos mensuráveis da vida dos indivíduos, largamente reconhecida como uma das maiores realizações da psicologia [7]. Estes traços são relativamente estáveis ao longo do tempo, mas diferem de indivíduo para indivíduo e influenciam o comportamento destes.

O modelo de Cinco Traços de Personalidade (*Big Five Personality Traits*) é produto de muitas décadas de pesquisa analítica, centrada nas características da personalidade humana. Foi originalmente concebido por [9] tendo suas

raízes nas hipóteses léxicas para identificação de traços de personalidade. A característica da abordagem léxica do modelo significa que os descritores de personalidade serão encontrados nas evidências da linguagem natural [11, 6].

As dimensões de personalidade identificadas neste modelo são: *Extraversion*, *Neuroticism*, *Agreeableness*, *Conscientiousness* e *Openness to Experience*. Cada uma destas dimensões do modelo *Big Five*, conhecidas pelo acrônimo OCEAN formado pelas suas letras iniciais, pode representar o perfil de um indivíduo em relação a esta dimensão específica. O estudo dos traços de personalidade está intimamente ligado ao processo de levantamento e medição, necessários para identificar as dimensões básicas da personalidade.

O pesquisador possui algumas hipóteses sobre o número e a natureza das principais dimensões, desenvolvendo um questionário para realizar a avaliação. Trabalhos posteriores investigaram a utilidade e validade do questionário e os adaptaram em função dos resultados obtidos [16].

O questionário *NEO Personality Inventory* (NEO-PI), desenvolvido por Costa e McCrae, demonstra excepcionais propriedades psicométricas sendo bem-sucedido em acomodar construtores já endereçados pelas métricas existentes de traços de personalidade. O termo NEO é um acrônimo formado com as iniciais das três dimensões inicialmente incluídas no estudo: *Neuroticism*, *Extraversion* e *Openness to Experience*. Inicialmente, Costa e McCrae incluíram escalas para medição de seis dimensões conceitualmente derivadas das três dimensões NEO, mas não incluíram as dimensões dos recentes fatores *Agreeableness* e *Conscientiousness*.

Com a publicação do NEO-PI-R: *NEO Personality Inventory - revised*, ampliaram o questionário para abranger 240 itens, incluindo estas duas novas dimensões. Ao contrário da maioria dos estudos léxicos da época, que eram baseados em amostras de estudantes de nível médio, o NEO-PI-R foi desenvolvido com amostras de estudantes de meia idade e adultos. Estas escalas demonstraram consistência substancial interna, estabilidade temporal, bem como convergência e validade discriminante [5].

ESTRUTURAÇÃO DE TEXTO

Um texto é por natureza uma informação primariamente não estruturada, ou seja, não é viável a realização de um processo de classificação de forma direta. Para isto, o texto precisa ser submetido a um processo de representação por meio do reconhecimento de características, para viabilizar a execução de técnicas de classificação, que permitam a identificação dos traços de personalidade do autor do texto. Os vetores de natureza numérica são mais adequados para a aplicação de técnicas de classificação. Estes vetores são obtidos utilizando diferentes estratégias que visam representar o texto, de acordo com a ótica do modelo adotado.

Um dos principais métodos utilizado para a obtenção de características do texto, é a utilização de léxicos, que

associam as palavras a certas categorias pré-definidas. Os léxicos permitem a categorização das palavras de acordo com algumas categorias, que podem incluir as emoções, no caso dos léxicos afetivos. A ferramenta LIWC (*Linguistic Inquiry and Word Count*), é uma ferramenta de software que permite o cálculo da frequência de ocorrência de diferentes categorias de palavras em um conjunto de textos fornecidos [23]. Esta ferramenta, tem sua funcionalidade baseada no dicionário LIWC, que é um léxico contendo cerca de 4.500 palavras previamente mapeadas em uma ou mais categorias diferentes, dentre as dezenas de categorias existentes neste dicionário.

Durante duas décadas, Dr. James Pennebaker realizou pesquisas sobre a relação entre linguagem, psicologia e saúde. Em conjunto com Martha Francis e Roger Booth, desenvolveu a ferramenta LIWC e o correspondente dicionário. Além do dicionário, no idioma inglês, a versão do LIWC2007 incorporou uma versão do dicionário em português [3].

Outros métodos, utilizam informações estatísticas sobre o texto. De uma forma simplista, pode ser realizado o levantamento da frequência de ocorrência de cada palavra no texto, e desta forma ser gerado um vetor que contenha esta informação. E este vetor é ampliado em função de novos textos que forem classificados na sequência.

Um destes métodos é o *n-gram*, que utiliza a predição de um próximo elemento (palavra) em uma sequência, na forma de uma ordem (n-1), no modelo probabilístico de Markov. Com isto, são obtidos os agrupamentos *unigram*, *bigram*, *trigram*. O agrupamento *unigram* tem como resultado, a frequência de ocorrência de cada palavra do texto. No caso do *bigram* têm-se as frequências de ocorrência dos pares consecutivos das palavras e o *trigram* adiciona as sequências de três palavras.

Supondo um texto que contenha 50 palavras diferentes, o tamanho do vetor *unigram* será 50, do *bigram* pode chegar a 2.500 e do *trigram* a 125.000. Outro método estatístico

disponível, é o *Word2Vec* [19], que utiliza o contexto linguístico das palavras, por meio da utilização de redes neurais superficiais. A partir de um conjunto de textos fornecidos este método produz um espaço vetorial com centenas de dimensões, associando para cada termo do *corpus* a um vetor numérico correspondente. Os vetores, são criados de modo que os termos que compartilham contextos comuns são posicionados próximos neste espaço vetorial.

MODELO IDENTIFICADOR DE PERFIL DE PERSONALIDADE PARAMETRIZÁVEL

O modelo apresentado por esta pesquisa, o Identificador de Perfil de Personalidade Parametrizável (IP3), consiste em um modelo de aprendizagem de máquina para classificação de texto que tem o objetivo realizar o reconhecimento dos traços de personalidade de textos de atividades educacionais, entre domínios diferentes.

A proposta está baseada no treinamento de um conjunto de classificadores, a partir de uma base de dados previamente identificada, que associa um conjunto de textos relacionados ao perfil OCEAN do autor de cada texto. Este conjunto de classificadores treinados é utilizado para o reconhecimento de um outro conjunto de dados, o texto dos alunos, que foi obtido a partir das atividades educacionais. A Figura 1 apresenta a arquitetura deste modelo.

O modelo apresentado tem como objetivo realizar a identificação das informações de traços de personalidade baseado somente no conteúdo do texto gerado pelos alunos, que no caso do experimento realizado, foi obtido a partir das informações armazenadas nas atividades de fórum de discussão. O modelo pode ser utilizado com outras fontes que contenham informação textual de autoria dos alunos oriundas do Ambiente Virtual de Aprendizagem (AVA) ou de outras fontes externas. As informações originadas no AVA são armazenadas em um arquivo no formato “CSV”, sendo os textos de cada aluno associados a um identificador

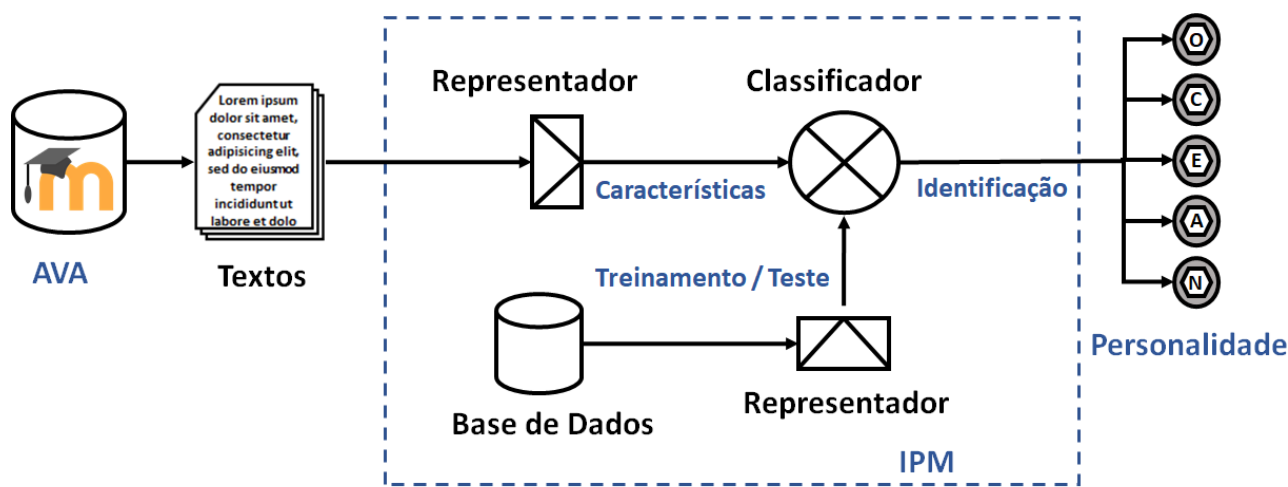


Figura 1. Modelo IP3

que representa o autor de cada texto. Os parágrafos do texto de cada autor, são agrupados sem os saltos de linha.

O Representador ilustrado na Figura 2 tem como função obter informações estruturadas sobre o texto dos alunos, para permitir a realização do processo de classificação. Foram utilizadas três formas de representação: léxico LIWC, *Word Embedding* e *Word2Vec*. No caso do LIWC, o texto original dos alunos foi utilizado para a estruturação, e nos demais, foi adicionada uma etapa intermediária de categorização (*POS tagging*), utilizando ferramentas de Processamento de Linguagem Natural (PLN). Esta categorização, permite obter a informação sobre a categoria sintática de cada palavra do texto, possibilitando a utilização da representação da categoria sintática das palavras.

Foi adotado o formato *Universal Part-of-Speech Tagset* [25], de forma a representar textos em diversos idiomas, sendo obtido um mesmo padrão de nome das categorias, que conforme este formato são: “NOUM”, “VERB”, “ADJ”, “ADV”, “PRON”, “DET”, “ADP”, “NUM”, “CONJ”, “PRT”, “.” e “X”.

O exemplo apresentado na Figura 3 ilustra o processo de categorização, utilizado uma frase como exemplo. A partir da frase de entrada “Personalidade é um conjunto de características que determinam os padrões de sentir e

proceder.”, fornecida como entrada para o módulo Categorizador, é obtido um vetor de termos correspondente a categoria de cada uma das palavras presentes nesta frase. A primeira palavra, “Personalidade”, é identificada como substantivo: (NOUN). A segunda palavra, “é”, corresponde a um verbo (VERB), e assim consecutivamente.

O extrator *Word Embedding* permite obter a representação da frequência de ocorrência dos “*n-gram*”, obtidos a partir das categorias sintáticas das palavras, de acordo com a técnica *Term Frequency–Inverse Document Frequency* (TF-IDF), gerando um conjunto de 5 arquivos SVM para *unigram*, 5 para *bigram* e 5 para *trigram*.

Da mesma forma, o extrator *Word2Vec* gera um conjunto de 5 arquivos contendo a representação obtida a partir da aplicação desta técnica nas categorias léxicas do conjunto de textos de entrada.

A partir da disponibilização destas diversas fontes de informação, que representam o texto estruturado, tanto da base de treinamento, quanto da base a ser identificada, o módulo Classificador realiza a inferência do perfil de personalidade dos alunos.

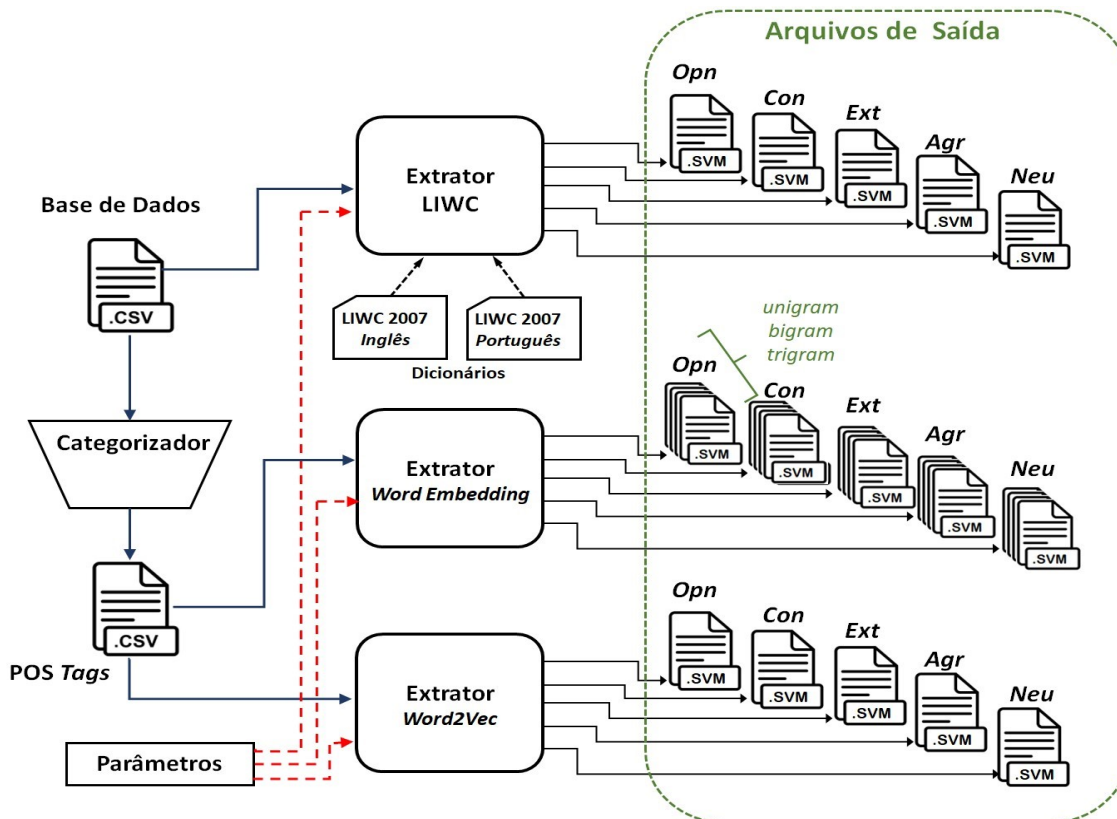


Figura 2. Representador.

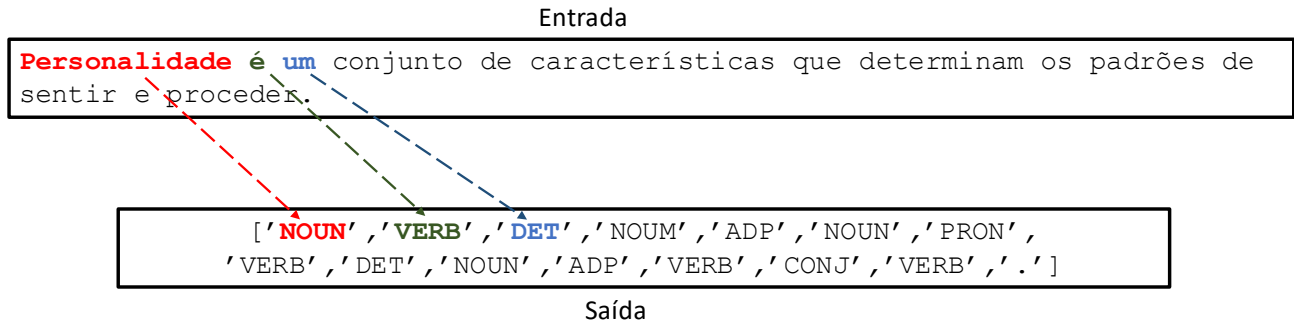


Figura 3. Exemplo de Categorização

São utilizados cinco classificadores, um para cada perfil do modelo *Big Five*, sendo que na Figura 4, está ilustrado o classificador para a dimensão *Extraversion*. Foi adotado o modelo *Ensemble* de classificação, onde um conjunto de classificadores é utilizado e o valor inferido ao final, é resultado do voto majoritário, obtido a partir dos resultados individuais de cada classificador.

Para cada classificador individual, pode ser configurada a forma de representação utilizada (LIWC, *Word2Vec*, *unigram*, ...) bem como o algoritmo de classificação utilizado (kNN, *Logistic Regression*, *Gaussian Naïve Bayes*,...). Desta forma, pode ser adotada, uma configuração específica para cada uma das dimensões do *Big Five*, em função de testes realizados com as diferentes bases e classificadores disponíveis, permitindo assim melhores resultados no processo de identificação do perfil de personalidade dos alunos.

EXPERIMENTO E RESULTADOS

Foi realizado um experimento em uma instituição brasileira de ensino superior, com três turmas de um curso na área de Educação. Em uma fase inicial, foi realizada uma breve apresentação dos objetivos da pesquisa aos professores responsáveis, sendo definido que seria realizada a aplicação de um questionário de levantamento de perfil de personalidade, a ser preenchido pelos alunos. O questionário utilizado foi uma versão do inventário BFI44 [12], adaptada e devidamente validada para o idioma português [1] por meio de uma análise fatorial, envolvendo uma amostra de 5.089 voluntários das cinco regiões brasileiras [27].

Após a coleta dos formulários preenchidos pelos 45 alunos, foi feita a compilação dos resultados e consequente obtenção manual do perfil de personalidade, para cada uma das dimensões do *Big Five*. Além destes indicadores, foram extraídos os conteúdos das mensagens registradas por estes alunos, nas atividades propostas pelos professores,

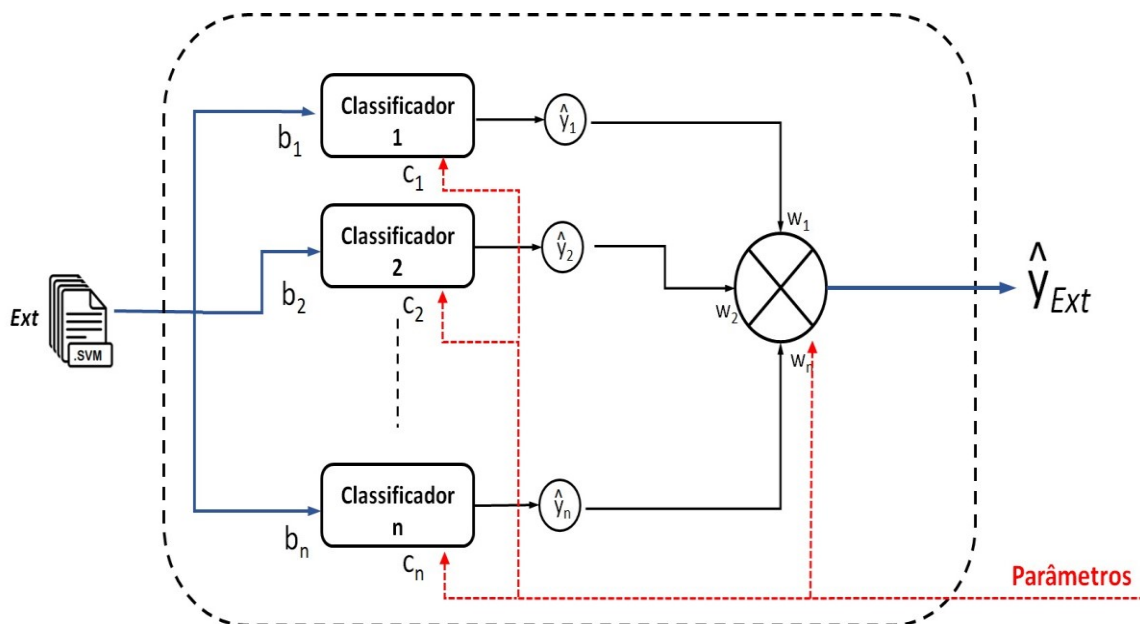


Figura 4. Classificador *Ensemble* para dimensão *Extraversion*.

utilizando a atividade de fórum de discussão, no ambiente Moodle. Este procedimento está ilustrado na Figura 5.

A obtenção do perfil de personalidade de forma manual, utilizando o formulário BFI44, foi realizada somente com o objetivo de realizar a validação inicial do modelo de identificação, sendo que em utilizações futuras deste modelo, somente será necessária a utilização dos textos dos alunos, para a realização desta identificação. Com os dados coletados, foi obtida uma base para validação do modelo, contendo o texto dos alunos e a identificação manual do perfil de personalidade, de acordo com o modelo *Big Five*, com os valores 0 e 1, correspondendo a presença ou ausência da dimensão específica.

Como base de treinamento, foi utilizada a base ESSAYS [23], em inglês, contendo dados coletados de 2.467 voluntários, composta pelo texto produzido por estes e pelo perfil de personalidade, identificado manualmente. Estas duas bases foram processadas no Representador do modelo IP3, obtendo um conjunto de arquivos SVM para serem utilizadas no processo de classificação.

A Tabela 1 apresenta os resultados da acurácia obtida utilizando o modelo IP3, com diversas combinações de

classificadores e técnicas de representação, no processo de classificação. Neste caso foi utilizada a base ESSAYS como base de treinamento e a base obtida junto aos alunos como base de teste. São apresentados também, os resultados verificados em experimentos similares, encontrados na literatura, que utilizam a base ESSAYS para validação dos modelos apresentados por estes pesquisadores.

O conjunto *Ensemble* utilizado no modelo IP3, teve configuração diferenciada para cada dimensão do *Big Five* no experimento realizado. Para a dimensão *Openness* foi utilizado a combinação dos classificadores *Random Forest* (RFOREST) e *Nearest Neighbors* (kNN) com a representação com LIWC, *Multi-layer Perceptron* (MLPC) com *unigram* e *Support Vector Machines* (SVM) com *Word2Vec*. No caso da *Conscientiousness* foi utilizada a combinação de 4 classificadores, sendo RFOREST e kNN com LIWC, MLPC com *unigram* e SVM com *Word2Vec*. Para *Extraversion* foram combinados 5 classificadores, *Logistic Regression* (LR), kNN e *Gaussian Naïve Bayes* (GNB) com representação LIWC, SVM com *unigram* e MLPC com *Word2Vec*. A dimensão *Agreeableness* utilizou um conjunto de 4 classificadores, sendo kNN, LR e RFOREST com

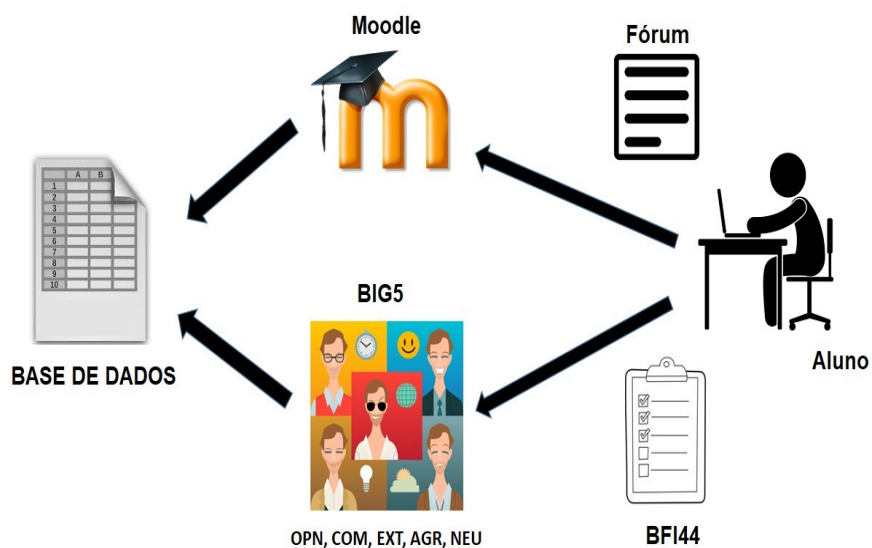


Figura 5. Obtenção da Base de Dados.

Modelo	OPN	CON	EXT	AGR	NEU
Mairesse (2006)	63%	56%	56%	56%	58%
Iacobelli (2013)	62%	55%	56%	53%	56%
Tighe (2016)	61%	55%	54%	57%	57%
Majumder (2017)	63%	57%	59%	57%	59%
Ensemble IP3	91%	62%	68%	66%	68%

Tabela 1. Acurácia obtida nos diversos experimentos.

LIWC e MLPC com representação *Word2Vec*. Finalmente para *Neuroticism* foram combinados os classificadores MLPC com LIWC e kNN, LR e MLPC com representação *unigram*.

Pode ser verificado que a utilização de uma combinação de técnicas de representação e classificadores apresenta resultados melhores do que a utilização de classificadores isolados. A utilização de *Word Embedding* ou *Word2Vec* isoladamente apresentou resultados inferiores do que a representação com LIWC. Mas, a combinação destas três técnicas, apresentou melhores resultados do que a utilização somente da representação LIWC nos processos de classificação.

CONSIDERAÇÕES FINAIS

Os resultados alcançados, demonstram a efetividade da utilização do modelo proposto para estimativa do Perfil de Personalidade, a partir das informações obtidas nos textos registrados pelos alunos, nas atividades educacionais. Na literatura pesquisada, não foram identificadas iniciativas de identificação de perfil de personalidade a partir de textos de atividades educacionais. Este tipo de texto tem características diferentes em relação aos textos livres produzidos pelos indivíduos em redes sociais onde a manifestação dos traços de personalidade é mais evidente. De outro modo, os valores encontrados, estão coerentes com os resultados obtidos por pesquisas que utilizaram textos oriundos de redes sociais e de outras fontes.

Dentre as técnicas de representação utilizadas, os melhores resultados individuais foram obtidos com o léxico LIWC. As técnicas de *n-gram* e *Word2Vec* apresentaram resultados inferiores isoladamente, mas quando utilizadas em conjunto com o LIWC, no *Ensemble* de classificadores, contribuíram para aumento da acurácia. O modelo apresentado, contribui para a área de educação, por apresentar uma alternativa automatizada para a identificação do perfil de personalidade dos alunos. Isto pode facilitar sobremaneira a utilização de estratégias educacionais personalizadas, nos ambientes de educação à distância.

Este modelo, oferece uma alternativa, para o problema da falta de bases consistentes, utilizadas no treinamento dos classificadores. Problema este, recorrente, nos trabalhos que tratam da identificação do perfil de personalidade, a partir de textos em português. Pelo modelo desenvolvido nesta pesquisa, que propõe a utilização dos dicionários léxicos em diversos idiomas e a categorização do texto antes da aplicação das técnicas de *Word Embedding* e *Word2Vec*, foi validada a eficácia da utilização da base de treinamento dos classificadores em idioma inglês, quando da identificação da personalidade utilizando textos em português, obtidos a partir das atividades dos alunos em Ambiente Virtual de Aprendizagem.

REFERENCES

1. Andrade, J. M. d. *Evidências de validade do inventário dos cinco grandes fatores de personalidade para o Brasil*. PhD thesis, Universidade de Brasília, 2008.
2. Argamon, S., Dhawle, S., Koppel, M., and Pennebaker, J. W. Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America* (2005).
3. Balage Filho, P. P., Pardo, T. A. S., and Aluísio, S. M. An Evaluation of the Brazilian portuguese LIWC Dictionary for Sentiment Analysis. *9th Brazilian Symposium in Information and Human Language Technology* (2013), 215–219.
4. Cohen, A., and Baruth, O. Personality, learning, and satisfaction in fully online academic courses. *Computers in Human Behavior* 72 (2017), 1–12.
5. Costa, P. T., and McCrae, R. R. *The NEO personality inventory*. PAR Psychological Assessment Resources, 1985.
6. De Raad, B. *The Big Five Personality Factors: The psycholexical approach to personality*. Hogrefe & Huber Publishers, 2000.
7. Deary, I. J. The trait approach to personality. *The Cambridge handbook of personality psychology* (2009), 89–109.
8. Ferreira, T., and Fernandes, M. Detecção de traços de personalidade em textos para apoiar a formação de grupos para colaboração. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (2017), vol. 28, p. 1627.
9. Galton, F. *The Measurement of Character*. Prentice-Hall, Inc, 1949.
10. Iacobelli, F., and Culotta, A. Too neurotic, not too friendly: structured personality classification on textual data. In *Proc of Workshop on Computational Personality Recognition*, AAAI Press, Melon Park, CA (2013), 19–22.
11. John, O. P., Angleitner, A., and Ostendorf, F. The lexical approach to personality: A historical review of trait taxonomic research. *European journal of Personality* 2, 3 (1988), 171–203.
12. John, O. P., and Srivastava, S. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* 2, 1999 (1999), 102–138.
13. Kassin, S. M. *Essentials of psychology*. Prentice Hall, 2003.
14. Mairesse, F., and Walker, M. Words mark the nerds: Computational models of personality recognition through language. *Proceedings of the 28th Annual*

- Conference of the Cognitive Science Society* (2006), 543–548.
15. Majumder, N., Poria, S., Gelbukh, A., and Cambria, E. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems* 32, 2 (2017), 74–79.
 16. Matthews, G., Deary, I. J., and Whiteman, M. C. *Personality traits*. Cambridge University Press, 2003.
 17. Melo, S., Dantas, A. C., and Fernandes, M. Modelo do estudante baseado em emoções e perfis de personalidade para recomendação de estratégias pedagógicas personalizadas. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (2017), vol. 28, p. 967.
 18. Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., and Reed, G. M. Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist* 56, 2 (2001), 128.
 19. Mikolov, T., Chen, K., Corrado, G., and Dean, J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013).
 20. Oberlander, J., and Nowson, S. Whose Thumb is It Anyway?: Classifying Author Personality from Weblog Text. In *Proceedings of the COLING/ACL on Main Conference Poster Sessions* (Stroudsburg, PA, USA, 2006), COLING-ACL '06, Association for Computational Linguistics, pp. 627–634.
 21. Omitido. In *Omitido durante processo de revisão* (2017).
 22. Omitido. *Omitido durante processo de revisão*. PhD thesis, Omitido, 2017.
 23. Pennebaker, J. W., Francis, M. E., and Booth, R. J. Linguistic inquiry and word count: LIWC 2007. *Mahway: Lawrence Erlbaum Associates* 71 (2001), 2001.
 24. Perrenoud, P. *A pedagogia na escola das diferenças: fragmentos de uma sociologia do fracasso*. Artmed, 2001.
 25. Petrov, S., Das, D., and McDonald, R. A universal part-of-speech tagset. arXiv preprint arXiv:1104.2086 (2011).
 26. Porto, S. M., Costa, W. S., Nunes, M., and Matos, L. N. Como a extração de personalidade através do teclado pode beneficiar a personalização na educação. *Anais do XXII SBIE-XVII WIE* (2011).
 27. Silva, B., and Paraboni, I. Learning personality traits from facebook text. *IEEE Latin America Transactions* 16, 4 (2018), 1256–1262.
 28. Tighe, E. P., Ureta, J. C., Pollo, B. A. L., Cheng, C. K., and de Dios Bulos, R. Personality trait classification of essays with the application of feature reduction. In *SAAIIP@IJCAI* (2016), pp. 22–28.
 29. Zheng, Y. Identifying dominators and followers in group decision making based on the personality traits. In *IUI Workshops* (2018).