

PARA ONDE CAMINHA A PESQUISA EM ENSINO DE FÍSICA? PERSPECTIVAS EVIDENCIADAS POR APRENDIZADO DE MÁQUINA

WHERE IS PHYSICS EDUCATION RESEARCH HEADING? PERSPECTIVES EVIDENCED BY MACHINE LEARNING

Erick Ghuron¹, Daniel Trugillo Martins Fontes², André Machado Rodrigues³

¹ Licenciando em Física da Universidade de São Paulo, ghuron@usp.br

² Programa de Pós-Graduação Interunidades em Ensino de Ciências da USP, daniel.fontes@usp.br

³ Instituto de Física da Universidade de São Paulo/Física Aplicada, rodrigues.am@usp.br

Resumo

Este trabalho explora tendências na pesquisa em Ensino de Física nos últimos vinte anos, com foco em trabalhos publicados do Encontro de Pesquisa em Ensino de Física (EPEF) entre 2002 e 2022. O objetivo é identificar as principais temáticas das publicações do EPEF e suas evoluções, utilizando o modelo de aprendizado de máquina *Latent Dirichlet Allocation* (LDA). Este algoritmo processou 1.958 trabalhos para detectar tópicos latentes, representando diferentes áreas de interesse e mudanças temáticas ao longo do tempo. Os resultados indicam temas predominantes como “Formação de Professores”, “Experimentação”, “Currículo” e “História da Ciência”, destacando a dinâmica e as prioridades da pesquisa na área. A análise sugere uma diversificação temática e mostra como a inteligência artificial pode auxiliar na compreensão das tendências em grandes volumes de dados. Conclui-se que o uso do LDA permitiu uma análise quantitativa das mudanças na pesquisa em Ensino de Física, oferecendo fundamentação para futuras investigações.

Palavras-chave: Revisão bibliográfica, Aprendizado de máquina, Tecnologia da informação e comunicação

Abstract

This work explores trends in Physics Education research over the last twenty years, focusing on papers published from the Encontro de Pesquisa em Ensino de Física (EPEF) between 2002 and 2022. The goal is to identify the main themes of the EPEF publications and their evolution, using the Latent Dirichlet Allocation (LDA) machine learning model. This algorithm processed 1,958 papers to detect latent topics, representing different areas of interest and thematic changes over time. The results indicate predominant themes such as “Teacher Training”, “Experimentation”, “Curriculum”, and “History of Science”, highlighting the dynamics and priorities of research in the field. The analysis suggests a thematic diversification and shows how artificial intelligence can assist in understanding trends in large volumes of data. It is concluded that the use of LDA allowed a quantitative analysis of the changes in Physics Education research, providing a foundation for future investigations.

Keywords: Literature Review, Machine Learning, Information and Communication Technology

Introdução ao problema de pesquisa

A pesquisa brasileira em Ensino de Física ocorre há cinco décadas, iniciando em meados da década de 1970 e desde 1980 realizando encontros científicos, possuindo programas de pós-graduação específicos, editando revistas científicas especializadas, entre tantas outras ações que identificam uma área de pesquisa madura (NARDI, 2005). Mais recentemente, com as várias edições do Encontro de Pesquisa em Ensino de Física (EPEF), surgiu e se estabeleceu a linha temática relacionada à metapesquisa ou questões teórico-metodológicas da pesquisa em Ensino de Física. Como tal linha temática é nomeada varia conforme a edição, mas podemos considerar que ela surgiu no V EPEF, em 1994, e se estabeleceu no X EPEF, em 2006. Atualmente ela está nomeada como *Questões teórico-metodológicas e novas demandas na pesquisa em Ensino de Física* que abarca, entre outros, a discussão acerca da identificação de tendências e perspectivas teóricas e metodológicas na pesquisa da área.

Ao mesmo tempo, precisamos reconhecer o aumento vertiginoso de publicações em Ensino de Física, e Ensino de Ciências no geral. Nesse contexto, trabalhos que pretendem investigar tendências precisam pensar em formas de avaliar várias centenas de trabalhos, superando desafios até pouco tempo atrás inéditos (FONTES e RODRIGUES, 2023). A leitura apenas do resumo por pesquisas que almejam traçar panoramas da área, por exemplo, traz limitações significativas (JACOMINI et al., 2023). Para atender a essa demanda, a literatura internacional em Ensino de Física e Ensino de Ciências tem sugerido e empregado o uso de inteligência artificial para auxiliar o trabalho do pesquisador na identificação de tendências ao analisar o texto completo de mais de 1.000 trabalhos (ver exemplos em CHEN et al., 2020; ODDEN, MARIN e CABALLERO, 2020; GHURON, FONTES e RODRIGUES, 2023). Diante do exposto, o presente trabalho se alinha a esse recente movimento, tendo como objetivo apontar quais são as tendências da pesquisa em Ensino de Física a partir da análise dos trabalhos publicados nos últimos vinte anos do EPEF. Para essa análise, desenvolvemos um modelo de aprendizado de máquina, conforme descrito a seguir.

Fundamentação teórica

O aprendizado de máquina é uma área da Inteligência Artificial, que ensina os computadores a aprenderem por meio de dados. Ao invés de programar o computador tradicionalmente, para fazer algo específico, ele é treinado com muitos exemplos de dados (MITCHELL, 1997).

Para esse trabalho utilizaremos o algoritmo *Latent Dirichlet Allocation* (LDA), que pertence à categoria dos algoritmos de aprendizado não supervisionado. Isso significa que nenhuma informação prévia sobre os textos é fornecida ao modelo. Ele também é um algoritmo probabilístico generativo que faz parte da classe dos modelos de tópicos, sendo uma forma de aprendizado de máquina que procuram descobrir os tópicos latentes em um conjunto de documentos (BLEI, NG e JORDAN, 2003). Neste modelo, cada documento é visto como uma combinação de vários tópicos, mas em proporções variadas. Ou seja, segundo o LDA, um documento não é associado a um único tópico, mas sim a uma distribuição deles. Cada tópico tem uma certa probabilidade de estar presente no documento, refletindo a frequência e a relevância das palavras associadas a esse tópico no texto, e cada tópico é representado como uma distribuição de palavras.

No modelo insere-se o número de tópicos que se deseja gerar e ajustam-se os hiperparâmetros α e η que controlam o grau de dispersão das distribuições de tópicos e palavras, respectivamente. Para avaliar a qualidade dos tópicos gerados, utiliza-se a métrica de *coerência*, tópicos que agrupam palavras frequentemente encontradas juntas em textos tendem a ser mais coerentes e, conseqüentemente, mais fáceis de interpretar (RÖDER, BOTH e HINNEBURG, 2015).

Caminhos da pesquisa

A obtenção dos trabalhos analisados foi realizada através da técnica *web scraping*, que consiste no processo de extrair informações presentes em sites da internet. Para isso utilizou-se um código de computador para automatizar o processo de: i) acessar os sites das edições anteriores do EPEF, ii) analisar o seu conteúdo, e iii) extrair os artigos desejados. No total foram coletados 1958 trabalhos, tanto pôster quanto comunicação oral, das 12 últimas edições do evento (de 2002 a 2022), por serem as únicas edições que possuem trabalhos disponíveis de forma online.

Após obter os trabalhos é necessário realizar um pré-processamento nos artigos. O primeiro passo foi extrair apenas os elementos textuais dos trabalhos, já

que o LDA é um modelo puramente textual (BLEI, NG e JORDAN, 2003). Com os textos extraídos realizamos outra limpeza, agora removendo palavras que não possuem semântica (e.g. artigos, pronomes, advérbios, etc) e ficamos apenas com as palavras que possuem significado ao texto, como: verbos, substantivos e adjetivos. Essa remoção é necessária, por ajudar a reduzir o volume de dados e a destacar as palavras que são mais significativas para a compreensão do conteúdo textual. Isso é particularmente útil na modelagem de tópicos, onde a identificação de termos relevantes é fundamental para descobrir os temas latentes no texto (GHURON, FONTES e RODRIGUES, 2023).

O segundo passo é a *lematização* das palavras restantes, isto é, a redução de palavras para suas *formas-base* ou *lemas*. Um lema é a forma dicionário de uma palavra, que representa seu significado independentemente das variações gramaticais (ODDEN, MARIN e CABALLERO, 2020). Ou seja, se for um verbo, se transforma para sua forma do infinitivo e se for um substantivo ou adjetivo transforma para sua forma do singular e masculina (caso o idioma em questão tenha gênero). Este processo permite a consolidação de diferentes formas de uma palavra, facilitando a análise e o entendimento do texto.

Após esse processo, filtramos as palavras restantes em relação a sua raridade ou generalidade. Essa filtragem garante que só possuímos palavras específicas para cada tópico. Com os dados processados, inserimo-nos no modelo LDA e escolhemos a quantidade de tópicos que o modelo irá gerar, para este trabalho escolhemos 9 tópicos. O modelo retorna a distribuição de palavras para cada tópico, com as primeiras palavras sendo as prováveis a pertencerem a esse tópico.

Resultados e discussão

O modelo apresentou bom resultado, com uma coerência de 0,51. Esse valor está na faixa esperada para trabalhos da área, que costumam ter coerências entre 0,4 e 0,6, conforme indicado por Syed e Spruit (2017). A **Tabela 1** mostra a distribuição de palavras para cada tópico latente e a quantidade de artigos que pertencem a cada tópico.

Tabela 1: Tópicos latentes dos trabalhos do Encontro de Pesquisa em Ensino de Física entre 2002 e 2022

Tópico	Palavras	Quantidade
Formação de Professores	profissional, licenciando, licenciatura, estágio, formação_professor, futuro, planejamento, universidade, formação_inicial, participante, futuro_professor, curricular, técnico, atuar, programa	280 (14,3%)
Experimentação	experimento, experimental, laboratório, atividade_experimental, ferramenta, simulação, computador, virtual, etapa, questionário, software, computacional, roteiro, tecnológico, prático	277 (14,1%)
Currículo	artigo, currículo, curricular, categoria, temática, livro_didático, evento, publicar, documento, temático, tecnológico, estado, selecionar, publicação, coleção	260 (13,3%)
História da Ciência	histórico, história, cientista, obra, cultura, cultural, humano, epistemológico, conhecimento_científico, pensamento, século, livro_didático, crítico, argumento, método	257 (13,1%)
Avaliação	item, questionário, astronomia, turma, prova, teste, desempenho, média, figura, motivação, alto, total, nota, tabela, índice	212 (10,8%)
Metodologia	professora, falar, achar, gente, jogo, coisa, criança, idéia, pergunta, episódio, explicação, série, fala, querer, linguagem	187 (9,5%)
Mecânica	movimento, velocidade, força, figura, luz, gráfico, corpo, onda, massa, lei, matemático, medida, som, posição, equação	172 (8,8%)
Física Térmica e Elétrica	energia, elétrico, calor, corrente, partícula, radiação, circuito, visual, temperatura, efeito, eletricidade, lâmpada, luz, mapa, magnético	170 (8,7%)
Argumentação	discurso, imagem, linguagem, enunciado, palavra, representação, gênero, divulgação, discursivo, rede, indivíduo, interdisciplinar, comunicação, homem, humano	143 (7,3%)

Com base nas palavras e nos artigos com maiores probabilidades de pertencer aos tópicos, possibilita-se nomeá-los com o tema em questão, como mostrado na tabela 1. Então, fazendo um gráfico da quantidade de artigos pelo intervalo de tempo, podemos analisar a evolução dos temas encontrados no EPEF.

Na **Figura 1** nota-se a evolução desses tópicos ao longo dos anos e a sua tendência linear. Observa-se que o tópico “Formação de professores”, apesar de ser o mais publicado, possui uma tendência de declínio. Essa tendência também foi notada por Salem (2012) ao analisar teses e dissertações sobre a pesquisa em Ensino de Física. O tópico “Experimentação” mostra uma tendência similar com mais altos e baixos, incluindo um pico notável em 2004, motivado pelo incentivo da comissão organizadora do EPEF IX a trabalhos dessa área¹. Vale notar um aumento

¹ Informação disponível em: < <https://sec.sbfisica.org.br/eventos/epef/ix/>>. Acesso em: 22 de abril 2024.

significativo de trabalhos publicados no tópico “Currículo” nos últimos anos, representando cerca de 30% das publicações no EPEF de 2022, crescimento também notado por Nurazmi e Bancong (2024) em contexto internacional.

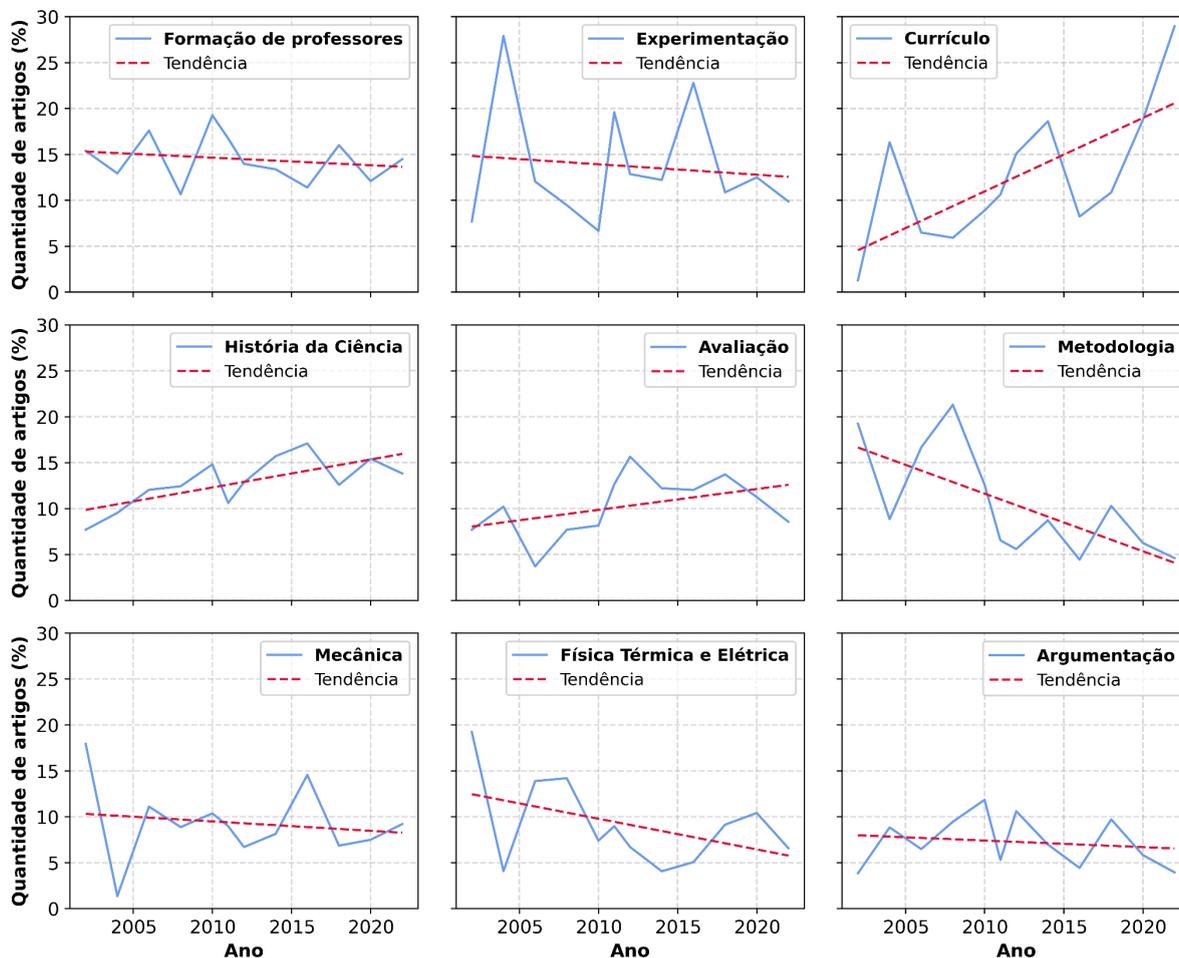


Figura 1: Frequência e tendência anual dos tópicos do Encontro de Pesquisa em Ensino de Física entre 2002 e 2022

O tópico “História da Ciência”, também aparece na análise de Salem (2012) com um crescimento ao longo do tempo, realizando trajetória semelhante ao tópico “Avaliação”, ambos com poucas variações acentuadas. Os tópicos de “Física Térmica e Elétrica” e “Mecânica” são tópicos especiais no sentido de que os trabalhos que estão nesses grupos debatem o Ensino de Física com essas disciplinas em perspectiva, e não que são pesquisas voltadas a essas áreas da Física.

Importante ressaltar que a pesquisa em Ensino de Física apresenta desafios significativos quando se trata de categorização e análise temática. A dificuldade

reside em desenvolver um modelo que não apenas agrupe palavras-chave eficazmente, mas também interprete a essência das publicações, refletindo as nuances e as diversidades da pesquisa em Ensino de Física, como nos tópicos “Metodologia” e “Argumentação”, que foram difíceis de nomear devido à subjetividade. A partir da análise dos artigos mais representativos identificados pelo modelo atual, percebe-se que uma parcela significativa deles se enquadra na categoria de *metapesquisas*. Este é um indicativo de que o modelo pode ser aprimorado para explorar melhor essa característica, permitindo uma compreensão dos temas e tendências predominantes na área. Portanto, um caminho possível e promissor para a evolução do modelo LDA é a incorporação de mecanismos que permitam uma análise mais refinada. Isso não apenas enriqueceria o entendimento da área na totalidade, mas também forneceria percepções valiosas para futuras investigações e direcionamentos na pesquisa em Ensino de Física.

Conclusões

Utilizando o LDA, foi possível identificar padrões e tópicos predominantes nos trabalhos, evidenciando a evolução e as prioridades da pesquisa na área. Os resultados obtidos com o modelo mostram a diversidade dos temas abordados no Ensino de Física. Estes tópicos refletem não apenas as áreas de interesse da comunidade acadêmica, mas também as necessidades e desafios enfrentados pelos educadores e pesquisadores na área. Nota-se a consistência do tópico “Formação de Professores”, indicando haver uma preocupação contínua de como preparamos os nossos futuros e presentes professores. A crescente importância do “Currículo” nos últimos anos aponta para uma reflexão sobre os conteúdos e as experiências pedagógicas adotadas, enquanto o interesse em “Física Térmica e Elétrica” sinaliza uma ênfase específica em certos domínios da Física. Apesar dos avanços proporcionados pelo LDA, ainda existem desafios a serem superados. A análise de metapesquisas e a captura da complexidade dos artigos demandam refinamentos no modelo. A habilidade de interpretar não apenas a frequência das palavras, mas também a essência dos textos, é necessária para uma compreensão mais aprofundada das tendências na pesquisa em Ensino de Física.

Por fim, este estudo contribui para uma melhor compreensão dos caminhos na pesquisa em Ensino de Física. As respostas obtidas através do LDA, foram

essenciais para acompanhar e interpretar a dinâmica da área, que está em constante evolução, e também oferecem um alicerce para futuras investigações.

Referências Bibliográficas

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent Dirichlet Allocation. **Journal of Machine Learning Research**, v. 3, p. 993–1022, 2003.

CHEN, X.; ZOU, D.; CHENG, G.; XIE, H. Detecting Latent Topics and Trends in Educational Technologies over Four Decades Using Structural Topic Modeling: A Retrospective of All Volumes of Computers & Education. **Computers & Education**, v. 151, p. 103855, 2020.

FONTES, D. T. M.; RODRIGUES, A. M. Sobre os modos de uso da teoria histórico-cultural na pesquisa em educação em ciências pelos membros dos Grupos de Pesquisa do CNPq. **Revista de Ensino, Educação e Ciências Humanas**, v. 24, n. 4, 2023.

GHURON, E.; FONTES, D. T. M.; RODRIGUES, A. M. Tópicos e tendências no ensino de física utilizando inteligência artificial. **Revista de Enseñanza de la Física**, v. 35, p. 167-173, 2023.

JACOMINI, M. A.; WELLEN, H. K. A. M.; PERRELLA, C. S. S.; MONÇÃO, M. A. G. Pesquisas estado da arte em educação: características e desafios. **Educação e Pesquisa**, v. 49, p. e262052, 2023.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.

NARDI, R. Memórias da educação em ciências no Brasil: a pesquisa em ensino de física. **Investigações em Ensino de Ciências**, v. 10, n. 1, p. 63-101, 2005.

NURAZMI, N.; BANCONG, H. Exploring physics education research: Popular topics in prestigious international journals in the period of 2009-2019. **AIP Conference Proceedings**, v. 2799, n. 1, p. 020126, 2024.

ODDEN, T. O. B.; MARIN, A.; CABALLERO, M. D. Thematic analysis of 18 years of physics education research conference proceedings using natural language processing. **Physical Review Physics Education Research**, v. 16, n. 1, 2020.

RÖDER, M.; BOTH, A.; HINNEBURG, A. **Exploring the Space of Topic Coherence Measures**. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. **Anais...** Em: WSDM 2015: EIGHTH ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING. Shanghai China: ACM, 2015.

SALEM, Sonia. **Perfil, evolução e perspectivas da pesquisa em ensino de física no Brasil. 2012**. Tese (Doutorado em Ensino de Física) - Ensino de Ciências (Física, Química e Biologia), Universidade de São Paulo, São Paulo, 2012.

SYED, S.; SPRUIT, M. **Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation**. 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). **Anais...** Em: 2017 IEEE INTERNATIONAL CONFERENCE ON DATA SCIENCE AND ADVANCED ANALYTICS (DSAA). Tokyo, Japan: IEEE, 2017.