



TENDÊNCIAS ATUAIS E FUTURAS EM PESQUISA DE MODELAGEM DE MUDANÇAS CLIMÁTICAS UTILIZANDO LDA E REGRESSÃO LINEAR

**SAMUEL FELIPE BUENO STEIMBACH GABRIELLA MARTINS NASCIMENTO
SKLAR**

RESUMO

O controle das mudanças climáticas é um tópico central de estudo devido à sua relevância para a sobrevivência da biodiversidade, saúde pública, segurança alimentar e outros aspectos importantes. Vários estudos utilizam modelos matemáticos ou computacionais para investigar as mudanças climáticas. Neste trabalho, foram coletadas publicações relevantes usando o banco de dados Scopus e, com o auxílio da técnica Latent Dirichlet Allocation (LDA), foi fornecido um panorama do cenário atual, destacando os estudos mais recorrentes relacionados às mudanças climáticas e modelagem. Além disso, foi realizada uma análise de regressão linear para prever tendências futuras nesses estudos. Os resultados indicaram que estudos focados na modelagem de mudanças climáticas em ambientes urbanos e no setor energético são atualmente os mais proeminentes. A modelagem em ambientes urbanos se destaca devido ao rápido crescimento das cidades e à intensificação dos impactos climáticos nessas regiões. O foco no setor energético reflete a necessidade de uma transição para fontes de energia sustentáveis, dado o papel central desse setor nas emissões de gases de efeito estufa. Esses resultados fornecem uma base importante para formular políticas públicas e direcionar pesquisas futuras, indicando áreas-chave para mitigação e adaptação às mudanças climáticas, com ênfase na modelagem como uma ferramenta essencial para entender e enfrentar esse desafio global.

Palavras-chave: Mudanças climáticas; Pesquisas; Alocação latente de Dirichlet; Modelagem climática; Tendências.

1 INTRODUÇÃO

As mudanças climáticas ocorrem naturalmente, mas as atividades humanas recentes têm exacerbado essas alterações (Pachauri e Reisinger, 2007). No último século, a temperatura aumentou em 0,74 °C, e a concentração de CO₂ subiu de 280 ppm para 368 ppm (Watson, 2001). As projeções indicam um aumento de temperatura de 3,4 °C e níveis de CO₂ atingindo 1250 ppm até 2095, juntamente com eventos climáticos extremos (Pachauri e Reisinger, 2007). A atividade industrial é a principal causa dessas mudanças (Watson, 2001). O Painel Intergovernamental sobre Mudanças Climáticas (IPCC) prevê que o aquecimento global pode em breve atingir 1,5 °C, destacando a urgência de ações climáticas para mitigar efeitos como perda de biodiversidade e insegurança alimentar (IPCC, 2023; Wang et. al, 2024). Ações imediatas podem melhorar a saúde pública e promover a justiça climática (IPCC, 2023).

Entender as mudanças do ecossistema em resposta às mudanças climáticas se tornou um ponto focal da pesquisa (McMahon et. al, 2011). Nesse contexto, a modelagem se tornou uma ferramenta crucial para quantificar e qualificar os impactos das alterações climáticas. Integrar a modelagem do ecossistema aos planos de ação climática é essencial para elaborar

estratégias que não apenas mitiguem os impactos climáticos, mas também aumentem a resiliência e as capacidades adaptativas dos ecossistemas, apoiando assim a saúde pública e promovendo a justiça climática (McMahon et. al, 2011).

Compreender tendências nas pesquisas de modelagem de mudanças climáticas, pode ser um instrumento importante. A alocação latente de Dirichlet (LDA), sugerida por Blei et al. (2003) como um algoritmo estatístico para descobrir tópicos potenciais em coleções de documentos extensas e atípicas, é a técnica de modelagem de tópicos mais simples e universalmente usada. A previsão de tendências usando modelagem de tópicos LDA surgiu como uma alternativa eficaz às técnicas tradicionais, como o método Delphi, que geralmente são caras e demoradas (Park et. al, 2018). A modelagem de tópicos LDA permite a descoberta de tópicos ocultos em vários documentos e realiza análises de previsão eficazes ao sugerir as proporções dos tópicos ocultos em todo o conjunto de documentos (Blei et al, 2003).

Integrar LDA em nossa análise nos permitirá identificar tendências emergentes e fornecer previsões robustas em estudos de modelagem de mudanças climáticas. LDA permite a análise de grandes coleções de documentos textuais, descobrindo tópicos latentes que ocorrem dentro de uma coleção de documentos e fornecendo uma representação explícita de cada documento com base em probabilidades de tópicos (Blei et al, 2003). Estudos anteriores conduzidos usando LDA incluem análises de tendências (Kin, 2016; Kin e Jang, 2016), estudos técnicos, como a extração de tópicos técnicos (Kin e Suh, 2017; Jeong, 2016).

O objetivo deste estudo é investigar as tendências atuais e futuras na pesquisa de modelagem de mudanças climáticas, utilizando técnicas como LDA e regressão linear, para identificar as áreas mais promissoras e fornecer previsões úteis para direcionamento de políticas públicas e novos estudos

2 MATERIAIS E MÉTODOS

Os dados para esta pesquisa foram coletados do banco de dados Scopus, usando palavras-chave relacionadas a mudanças climáticas e modelagem matemática ou computacional. O Scopus foi escolhido devido à sua ampla gama de periódicos de pesquisa (Burnham, 2006). As palavras-chave foram estruturadas na consulta de pesquisa avançada do Scopus da seguinte forma: (TITLE ("climate change") OR TITLE ("global warming") OR TITLE ("climate model") OR TITLE ("climate modelling") OR TITLE ("climate simulation") OR TITLE ("climate forecast")) AND (TITLE("modeling") OR TITLE("modelling") OR TITLE("simulation") OR TITLE("forecasting") OR TITLE("prediction") OR TITLE("projection")). O período de tempo considerado foi de 1974 a 2024, resultando em um total de 9.567 artigos.

O pré-processamento dos resumos envolveu várias etapas para garantir que os dados fossem adequados para análise. Iniciou-se com a remoção de palavras irrelevantes que não contribuem com significado significativo para o texto (por exemplo, 'and', 'the', 'of') foram removidas usando o parâmetro *stop_words* definido como 'english' no *TfidfVectorizer* da biblioteca scikit-learn (Pedregosa et. al, 2011). Em seguida os textos foram transformados em vetores numéricos usando a técnica Term Frequency-Inverse Document Frequency (TF-IDF). Isso foi implementado com o *TfidfVectorizer* da biblioteca scikit-learn. O TF-IDF ajuda a identificar as palavras mais importantes em cada documento, tornando os dados mais informativos para análise posterior. O parâmetro *max_features* foi definido como 1000 para limitar o número de recursos (Pedregosa et. al, 2011).

2.2. Modelagem de Tópicos

A análise de tópicos foi conduzida usando o modelo Latent Dirichlet Allocation (LDA), implementado com a biblioteca scikit-learn. LDA é um modelo probabilístico generativo para coleções de dados discretos, como corpora de texto, conforme descrito por

Blei et al. (2003). O modelo LDA foi treinado nos vetores TF-IDF. Isso envolveu a especificação do número de tópicos (definido como 10) e o uso de um estado aleatório para reprodutibilidade. O parâmetro *n_components* foi definido como 10 para identificar dez tópicos principais, e *random_state* foi definido como 42 para reprodutibilidade. Cada tópico foi representado por um conjunto de palavras-chave mais frequentes. Essa técnica permite a descoberta de tópicos ocultos em um grande conjunto de documentos, facilitando a identificação de padrões e tendências (Park et. al, 2018) na pesquisa sobre mudanças climáticas.

2.4. Previsão de Tendências com Modelo de Regressão Linear

Para prever as tendências futuras de publicações, empregamos um modelo de Regressão Linear. A Regressão Linear visa encontrar a relação linear entre a variável dependente (número de publicações) e a variável independente (ano). O modelo de Regressão Linear foi treinado usando o número agregado anual de publicações para cada tópico. A classe *LinearRegression* do scikit-learn foi usada para esse propósito. Usando o modelo treinado, projetamos o número de publicações para cada tópico de 2024 a 2030. Isso envolveu ajustar o modelo aos dados existentes e então usá-lo para prever valores futuros (Pedregosa et. al, 2011). Para quantificar a incerteza das previsões, calculamos o erro padrão e os intervalos de confiança para as previsões. O erro padrão foi computado usando os resíduos do modelo, e os intervalos de confiança foram derivados com base na distribuição normal padrão (Pedregosa et. al, 2011).

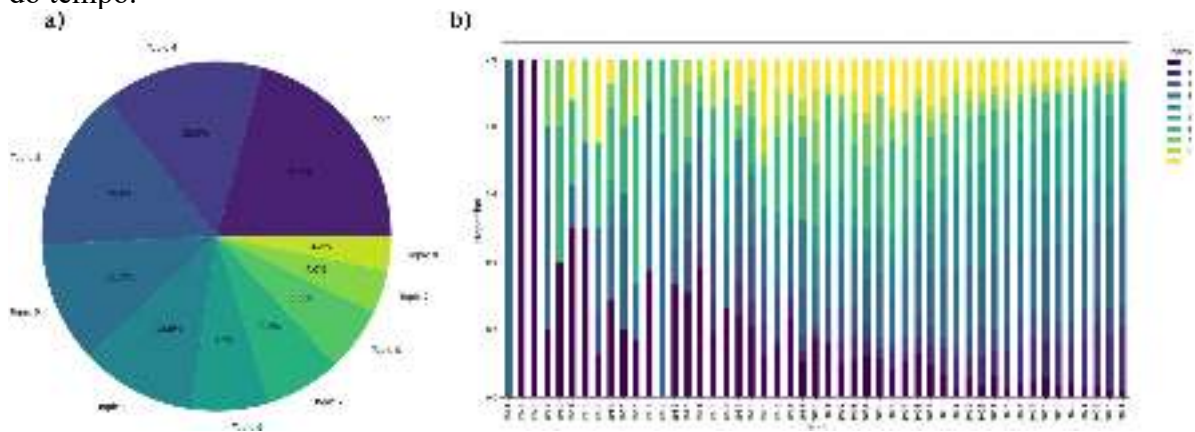
Na abordagem de validação cruzada usada no contexto de previsão de tópicos de publicação científica, a técnica de validação cruzada *k-fold* foi aplicada para validar a robustez do modelo de regressão linear. Essa abordagem ajuda a garantir que o modelo se generalize bem para novos dados e não esteja superajustando os dados de treinamento. Os dados históricos de publicação foram divididos em *k folds* (ou dobras). Em cada iteração, uma dobra foi usada como o conjunto de validação, enquanto as *k-1 folds* restantes foram usadas como o conjunto de treinamento (Pedregosa et. al, 2011). Uma validação cruzada de 10 *folds* foi usada, ou seja, os dados foram divididos em 10 partes.

3 RESULTADOS E DISCUSSÃO

As análises realizadas após o pré-processamento dos dados usando LDA escolheu-se 10 tópicos para o modelo. Cada registro bibliográfico dos dados coletados foi atribuído a um dos dez tópicos. O conjunto de dados resultante dessa classificação permitiu uma análise detalhada dos tópicos ao longo do tempo.

Na Figura 1, no gráfico a), podemos observar as porcentagens ocupadas pelas classes na amostra total, com os tópicos 3 e 4 sendo mais relevantes na análise geral. No gráfico b), podemos observar as variações das classes ao longo de 50 anos (1974-2024), e na Tabela A1, é possível ver sua frequência e variação. Nesse contexto, o tópico 3 apresentou o maior crescimento ao longo do tempo (para entender suas palavras-chave e a relação entre as palavras-chave, consulte a Tabela 1). Além disso, a Tabela A1 fornece informações adicionais sobre a análise histórica.

Figura 1 - Gráficos, onde a) é a proporção de cada um dos tópicos e b) sua variação ao longo do tempo.



Fonte: Autores (2024)

Tabela 1 - Tabela de tópicos e palavras-chave.

Tópico	Palavras-chaves	Possível Relação
Topic #0	clouds, ice, climate, transition, snow, scheme, land, model, surface, cloud	Concentra-se na modelagem de fenômenos atmosféricos e de superfície, como nuvens e gelo.
Topic #1	species, article, mean, published, 2023, error, reproducibility, forcing, available, abstract	Enfatiza a publicação de artigos e a disponibilidade de dados, especialmente nos regimes árticos.
Topic #2	maize, rice, yields, water, wheat, production, change, climate, yield, crop	Concentra-se nos impactos das mudanças climáticas na agricultura, como na produtividade das colheitas e na produção agrícola.
Topic #3	research, paper, urban, modeling, model, energy, data, models, change, climate	Foca na modelagem de mudanças climáticas em ambientes urbanos e energia.
Topic #4	temperature, bias, rom, resolution, models, simulations, regional, model, precipitation, climate	Concentra-se em modelagem e simulações climáticas regionais, abordando resolução e vieses de modelos regionais.
Topic #5	scenario, hydrological, incision, basin, change, runoff, future, river, climate, water	Foca em modelagem hidrológica e cenários futuros de precipitação e escoamento.
Topic #6	surface, temperature, global, ocean, changes, model, climate, warming, sea, ice	Enfatiza os aspectos globais das mudanças climáticas, incluindo o aquecimento global e as mudanças nos oceanos.
Topic #7	species, article, mean, published, 2023, error, reproducibility, forcing, available, abstract	Destaca a importância da disponibilidade de dados e da reprodutibilidade de modelos, abordando erros comuns na modelagem.
Topic #8	change, app, lake, climate, drought, recharge, vegetation, water, soil, groundwater	Concentra-se na interação entre mudanças climáticas e recursos hídricos, incluindo vegetação e recarga de águas subterrâneas.
Topic #9	variability, model, climate, ocean, time, predict, ice, precipitation, model, namaste	Concentra-se na variabilidade climática e nos fenômenos oceânicos, como ENSO e monções.

Fonte: Autores (2024)

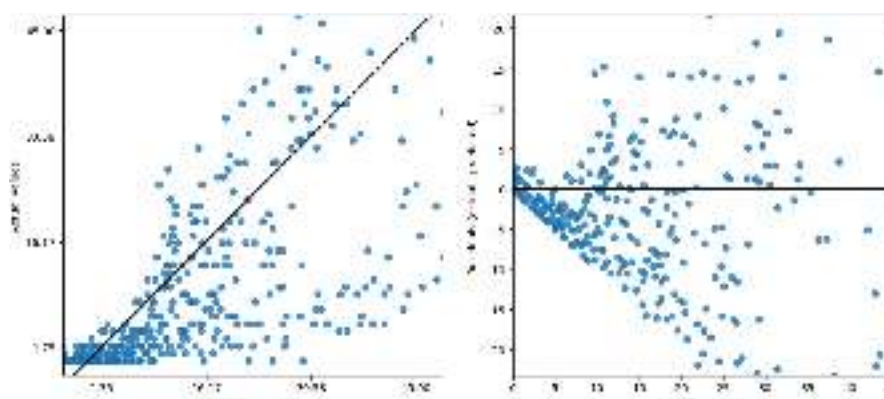
Após analisar tendências históricas, realizamos regressão linear com coeficientes para minimizar a soma dos quadrados residuais entre os valores observados no conjunto de dados e os valores previstos pela aproximação linear, visualizada na Figura 3.

As previsões quando comparadas aos valores reais, ou seja, aqueles da tendência histórica, analisados até os dados de 2024, no eixo y, com os valores previstos, com a previsão do ano de 2024 a 2030, no eixo x (Figura 2., sem gráficos à esquerda). Idealmente, eles devem se alinhar com a linha preta pontilhada diagonal, indicando desvios perfeitos. Há uma concentração significativa de pontos na parte inferior dos gráficos, o que pode indicar um baixo número de publicações em muitos tópicos. À medida que a contagem de publicações

aumenta, um aumento na dispersão pode ser observado, com ênfase menos precisa em valores maiores e detalhados de publicações.

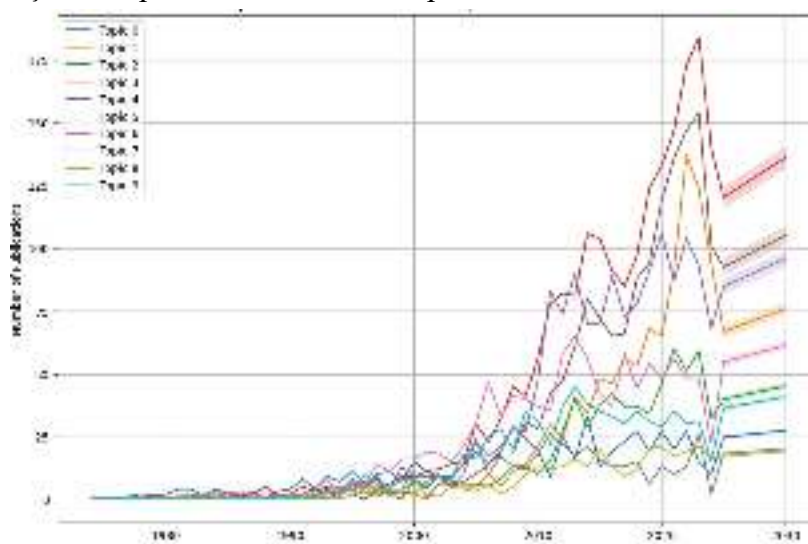
No gráfico à direita (Figura 2.), observamos no eixo y, os resíduos, ou seja, a diferença entre os valores reais e previstos, e no eixo x, os valores previstos. A presença de técnicas de detalhe no eixo horizontal, ou seja, Residual = 0, indica que muitos tópicos estão próximos dos valores reais (ou seja, aquela interpretação da tendência histórica). Assim como o outro gráfico (o da esquerda), este também aumenta o número de publicações, resultando em menor precisão para tópicos com valores mais altos.

Figura 2 - Plotagem das previsões realizadas cruzadamente para tópicos, onde à esquerda está a comparação dos valores atuais e das previsões. E à direita está a comparação dos valores residuais com os valores previstos.



Fonte: Autores (2024)

Figura 3 - Evolução Temporal e Previsão de Tópicos com Intervalo de Confiança



Fonte: Autores (2024)

No gráfico da Figura 3, é possível observar a variação das tendências ao longo dos 50 anos (1974 a 2024), mais as tendências previstas para os anos de 2024 a 2030, com seus respectivos intervalos de confiança. Essa representação gráfica, em valores, está descrita na Tabela 2.

Os resultados obtidos neste estudo destacam tendências significativas nas pesquisas de modelagem de mudanças climáticas, com ênfase especial em tópicos relacionados a ambientes urbanos e ao setor energético. A predominância destes temas na literatura pode ser

interpretada como um reflexo do impacto crescente da urbanização rápida e da necessidade urgente de transição para fontes de energia sustentável, alinhando-se com estudos anteriores que indicam a vulnerabilidade particular de áreas urbanas e a contribuição do setor energético para as emissões de gases de efeito estufa (McMahon et al., 2011). Esse foco indica uma demanda para o desenvolvimento de modelos que abordem de forma mais eficaz os efeitos específicos das mudanças climáticas em áreas densamente povoadas e em setores de alta emissão.

No que diz respeito à modelagem de impacto em cultivos agrícolas e mudanças hidrológicas, os resultados apontam para uma preocupação crescente com a segurança alimentar e a gestão da água, duas áreas cruciais para enfrentar as consequências das mudanças climáticas globais (Watson, 2001). As palavras-chave associadas a esses tópicos indicam uma tendência em compreender como variáveis climáticas, como temperatura e precipitação, influenciam a produtividade agrícola e a disponibilidade de recursos hídricos, especialmente em bacias hidrográficas sensíveis. Essas descobertas reforçam a importância de estudos que integram modelagem climática e análise de impacto ambiental, uma abordagem defendida por Pachauri e Reisinger (2007), que propõem modelos mais dinâmicos para capturar a complexidade dos ecossistemas em mudança.

4 CONCLUSÃO

A modelagem matemática e computacional é uma ferramenta fundamental para o estudo das mudanças climáticas globais. Em linha com essa preocupação, um sistema de previsão de tendências foi desenvolvido usando a técnica de modelagem de tópicos Latent Dirichlet Allocation (LDA). Primeiro, um framework foi criado para estruturar o mecanismo de previsão. Em seguida, os fluxos de trabalho do sistema de modelagem de tópicos LDA e do sistema de análise de tendências em estudos envolvendo mudanças climáticas e modelagem foram detalhados. A análise identificou tendências em dez tópicos principais na pesquisa de modelagem de mudanças climáticas, com ênfase em áreas como modelagem em ambientes urbanos e setor de energia, modelagem regional, cenários hidrológicos e mudanças oceânicas. A partir da análise histórica, observamos um aumento exponencial nas publicações desde a década de 1990, com crescimento contínuo esperado nos próximos seis anos. As previsões indicam um crescimento especialmente acentuado em tópicos que já estão bem estabelecidos. Finalmente, os resultados obtidos da avaliação dos modelos de previsão usando regressão linear podem indicar que muitos tópicos gerados pelo LDA estão próximos dos valores verdadeiros, confirmando a precisão do modelo para a maioria dos tópicos analisados. No entanto, para tópicos com um número maior de publicações, a precisão diminuiu, refletindo a crescente complexidade desses tópicos. As implicações práticas deste estudo são significativas: o sistema de previsão desenvolvido oferece uma ferramenta eficiente, rápida e de baixo custo para análise de tendências em pesquisa climática. As limitações deste estudo incluem a necessidade de explorar métodos estatísticos adicionais para análise de tendências que sejam mais rápidos, além da regressão linear. Pesquisas futuras devem investigar métodos de verificação mais eficazes e considerar a implementação prática dos sistemas de previsão desenvolvidos.

Em resumo, este estudo contribui para a compreensão das direções futuras na pesquisa sobre mudanças climáticas, fornecendo uma base sólida para o desenvolvimento de estratégias de pesquisa e políticas públicas. Identificar tendências emergentes e prever desenvolvimentos futuros são fundamentais para direcionar esforços e recursos de pesquisa estrategicamente, contribuindo para uma resposta mais eficaz às mudanças climáticas globais.

REFERÊNCIAS

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

BURNHAM, J. F. **Scopus database: a review**. Biomed Digit Libr, v. 3, n. 1, 2006.
Disponível em: <https://doi.org/10.1186/1742-5581-3-1>.

IPCC. **Climate Change 2023: The Physical Science Basis**. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 2023.

JEONG, B. K.; LEE, H. Y. **Research Topics in Industrial Engineering 2001~2015**. Journal of the Korean Institute of Industrial Engineers, v. 42, n. 6, p. 421-431, 2016.

KIM, J. H. **Research Trends Analysis for ‘Internet of Things’ Based on Topic Modeling and Network Analysis**. 2016. Dissertação (Mestrado) – Seoul National University of Science and Technology, 2016.

KIM, S. K.; JANG, S. Y. **A Study on the Research Trends in Domestic Industrial and Management Engineer**. Journal of the Korea Management Engineers Society, v. 21, n. 3, p. 71-95, 2016.

KIM, M. A.; SUH, C. K. **SCM Patent Analysis Using Topic Modeling: 1997~2016**. Journal of the Korean Society of Supply Chain Management, v. 17, n. 2, p. 19-29, 2017.

MCMAHON, S. M. et al. **Improving assessment and modelling of climate change impacts on global terrestrial biodiversity**. Trends in Ecology and Evolution, v. 26, n. 5, p. 249-259, 2011. Disponível em: <https://doi.org/10.1016/j.tree.2011.02.012>.

PACHAURI, R. K.; REISINGER, A. **Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change**. IPCC, 2007.

PARK, J. S. et al. **A New Forecasting System using the Latent Dirichlet Allocation (LDA) Topic Modeling Technique**. WSEAS Transactions on Environment and Development, v. 14, p. 363-372, 2018.

PEDREGOSA, F. et al. **Scikit-learn: Machine Learning in Python**. Journal of Machine Learning Research, v. 12, p. 2825-2830, 2011.

WANG, M. et al. **A triple increase in global river basins with water scarcity due to future pollution**. Nature Communications, v. 15, p. 880, 2024.

WATSON, R. T. **Climate Change 2001: Synthesis Report**. IPCC, 2001.