



Previsão de deslocamento químico de complexos de Pt195 por meio de Machine Learning

Daniel P. PEREIRA¹; Artur G. NOGUEIRA²; Ricardo R. de OLIVEIRA³

RESUMO

Este artigo explora a previsão de deslocamento químico para compostos de coordenação de platina, com foco em sua estrutura eletrônica, essencial para aplicações em catálise, medicina e ciência dos materiais. A análise do deslocamento químico fornece informações valiosas sobre a estereoquímica e reatividade desses compostos. A fundamentação teórica abrange a importância de considerar efeitos relativísticos em cálculos envolvendo elementos pesados, como a platina. O estudo propõe a utilização de técnicas de Machine Learning para prever o deslocamento químico, visando otimizar recursos e tempo em cálculos complexos. Utilizando um conjunto de dados construído a partir de compostos previamente estudados, algoritmos como Regressão Linear e Random Forest foram treinados com atributos específicos dos compostos. Surpreendentemente, a Regressão Linear apresentou resultados altamente satisfatórios, com acurácia de 90-95% e RMSE de 272, superando as expectativas em relação à necessidade de redes neurais.

Palavras-chave: Compostos complexos de Platina; Machine Learning e RMN

1. INTRODUÇÃO

O estudo de previsão de deslocamento químico para os compostos de coordenação de platina pode fornecer informações valiosas sobre a estrutura eletrônica destes compostos. Esses compostos são amplamente utilizados em muitas áreas, incluindo a catálise (TEHFE et al., 2012) (ABE et al., 2001), a medicina (LIPPERT; SANZ MIGUEL, 2016) (SUN et al., 2021) (KRITCHENKOV et al., 2019) e a ciência dos materiais (LIEN et al., 2020) (TANASE et al., 2019). Um dos principais fatores que determinam as propriedades destes compostos é a estrutura eletrônica do átomo de platina e sua interação com os ligantes circundantes. O deslocamento químico é uma poderosa ferramenta para analisar a estrutura eletrônica dos compostos de coordenação de platina (GOGGIN et al., 1976). O cálculo do deslocamento químico também pode ser utilizado para identificar o ambiente de coordenação que compõe o composto estudado e sua estereoquímica, características frequentemente cruciais para compreender a reatividade, atividade catalítica e propriedades conformacionais destes compostos (BAJAJ et al., 2018).

2. FUNDAMENTAÇÃO TEÓRICA

Em geral, a previsão do deslocamento químico, dos efeitos relativistas e da solvatação exige uma combinação de abordagens teóricas e experimentais, bem como um conhecimento profundo dos princípios químicos subjacentes e dos métodos computacionais. Como concluiu Castro (2010):

¹Bolsista PIBIC/CNPq, IFSULDEMINAS – *Campus* Poços de Caldas. E-mail: daniel.pereira@alunos.ifsulde Minas.edu.br.

²Pós Graduando, UFLA – *Campus* Lavras. E-mail: artur.nogueira1@estudante.ufla.br.

³Orientador, IFSULDEMINAS – *Campus* Poços de Caldas. E-mail: ricardo.ramos@ifsulde Minas.edu.br.

qualquer predição de deslocalização de elétrons em moléculas contendo elementos pesados sem considerar um tratamento adequado para efeitos relativísticos pode levar a uma interpretação química errônea. A implementação dos efeitos relativísticos torna-se um importante desafio a ser superado na química computacional no intuito de prever valores de deslocamento químico, como defendeu Demissie (2013). Quando se trata de predição, a área de Machine Learning tem como uma de suas premissas poupar tempo e recursos que seriam gastos com técnicas complexas. De maneira especial, a predição do deslocamento químico de moléculas é um prospecto para as técnicas de aprendizado de máquina, pois, como afirma Carleo (2019), métodos de Machine Learning são particularmente úteis quando muitos cálculos são requeridos, como no caso do cálculo do deslocamento químico. O mesmo é afirmado por Wei (2019): machine learning pode acelerar o processo de predição das propriedades de materiais, o que tipicamente exigiria cálculos computacionais dispendiosos.

3. MATERIAL E MÉTODOS

A base de dados utilizada consistia inicialmente em uma planilha com a descrição detalhada das propriedades de várias moléculas de platina, e foi dividida em duas partes, uma com aproximadamente oitenta por cento das moléculas, servindo para treinamento, e outra menor com os vinte por cento restantes, para teste. O conjunto baseou-se no artigo de Paschoal (2016), no qual reuniu nove grupos de Compostos de Coordenação.

O trabalho desenvolvido com machine learning começou com um processo de análise e estudo das diversas moléculas de platina e suas características, por meio de técnicas de estatística disponibilizadas pelas bibliotecas da linguagem Python, como Pandas, Numpy e Matplotlib, pelas quais foram obtidas informações valiosas que serviram para a melhor modelagem da base de dados e a escolha de quais algoritmos seriam inicialmente utilizados. Após isso, a biblioteca Scikit-learn, também provinda da linguagem Python e focada em técnicas de aprendizado de máquina, foi utilizada para a criação de diferentes modelos de algoritmos de previsão de dados, que receberam como entrada o contingente da base de dados destinado ao treinamento destes. Por último, os modelos foram aplicados ao conjunto de dados de teste, e os resultados obtidos foram comparados com os valores reais e também os de outros modelos.

4. RESULTADOS E DISCUSSÃO

A expectativa inicial era de que algoritmos mais simples de aprendizado de máquina como Regressão Linear e Random Forest fossem ter um desempenho abaixo e que seria necessário modelos mais robustos como uma rede neural, devido a semelhança da representação de uma molécula com um grafo, para a obtenção de melhores resultados. No entanto, os testes com rede

neural ainda estão começando a serem desenvolvidos, pois os resultados obtidos com o modelo de regressão linear foram positivamente surpreendentes e por conta disso um tempo extra foi gasto utilizando técnicas de validação para garantir que estes resultados não eram enganosos. Felizmente, foram obtidas diversas comprovações de que a performance do modelo de regressão era consistente e válida. Os números obtidos foram entre 90 e 95 por cento de acurácia para a medida estatística r-quadrado e 272 para o RMSE.

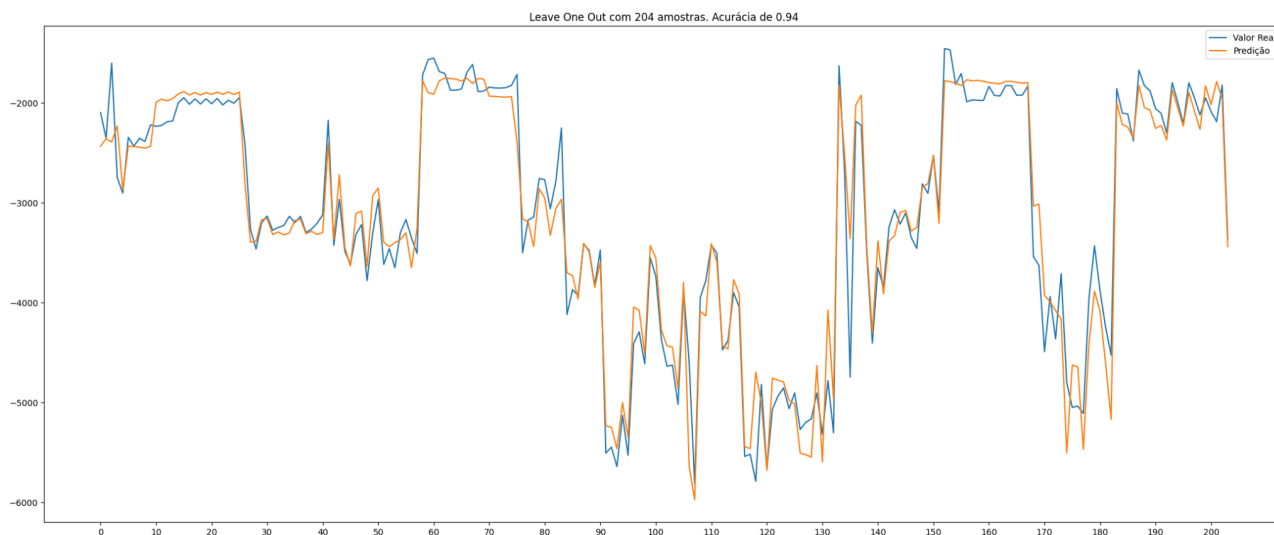


Figura 1: Comparação entre o valor real e o predito pelo modelo de regressão linear, utilizando a técnica de validação cruzada Leave One Out, com acurácia de 94%

5. CONCLUSÃO

Os resultados obtidos com o modelo de Regressão Linear demonstraram ser promissores e eficientes, superando as expectativas iniciais e se mostrando adequados para a previsão de deslocamento químico em Compostos de Coordenação de Platina. Embora o estudo com redes neurais ainda esteja em fase inicial, a alta precisão dos modelos de Regressão Linear sugere que, com validação adequada, esses métodos podem ser uma solução prática para prever propriedades moleculares, sem a necessidade imediata de modelos mais complexos. O estudo, assim, contribui para a aplicação de técnicas de Machine Learning na Química Computacional, otimizando a análise de compostos de elementos pesados e abrindo portas para futuras investigações que possam integrar algoritmos mais robustos para melhoria contínua.

REFERÊNCIAS

ABE, T. et al. Electrocatalysis for proton reduction by polypyridyl platinum complexes dispersed in a polymer membrane. **European Polymer Journal**, v. 37, n. 4, 2001.

BAJAJ, Naini; BHATT, Himat; PANDEY, K. K.; POSWAL, H. K.; ARYA, A.; GHOSH, P. S.; GARG, N.; DEO, M. N.. Phase transition in metal–organic complex trans-PtCl₂(PEt₃)₂ under

pressure: insights into the molecular and crystal structure. **Crystengcomm**, [S.L.], v. 20, n. 26, p. 3728-3740, 2018. Royal Society of Chemistry (RSC).

CARLEO, G. et al. Machine learning and the physical sciences. **Reviews of Modern Physics**, n. 91, 2019.

CASTRO, A. C. et al. Scalar and spin-orbit relativistic corrections to the NICS and the induced magnetic field: The case of the E122- spherenes (E = Ge, Sn, Pb). **Journal of Chemical Theory and Computation**, v. 6, n. 9, 2010.

DEMISSIE, T. B. et al. Four-component relativistic chemical shift calculations of halogenated organic compounds. **Journal of Physical Organic Chemistry**, v. 26, n. 8, 2013.

GOGGIN, Peter L.; GOODFELLOW, Robin J.; HADDOCK, Steve R.; TAYLOR, Brian F.; MARSHALL, Iain R. H.. The platinum-195 chemical shift in some platinum(0) and platinum(II) complexes and its relationship to their structure. **Journal Of The Chemical Society, Dalton Transactions**, [S.L.], n. 5, p. 459, 1976. Royal Society of Chemistry (RSC).

KRITCHENKOV, A. S.; STANISHEVSKII, Ya. M.; SKORIK, Yu. A.. Design and Antitumor Activity of Platinum Complexes. **Pharmaceutical Chemistry Journal**, [S.L.], v. 53, n. 1, p. 6-14, 13 abr. 2019. Springer Science and Business Media LLC.

LIEN, Chen-Yu; HSU, Ying-Feng; LIU, Yi-Hung; PENG, Shie-Ming; SHINMYOZU, Teruo; YANG, Jye-Shane. Steric Engineering of Cyclometalated Pt(II) Complexes toward High-Contrast Monomer–Excimer-Based Mechanochromic and Vapochromic Luminescence. **Inorganic Chemistry**, [S.L.], v. 59, n. 16, p. 11584-11594, 3 ago. 2020. American Chemical Society (ACS).

LIPPERT, B.; SANZ MIGUEL, P. J. More of a misunderstanding than a real mismatch? Platinum and its affinity for aqua, hydroxido, and oxido ligands. **Coordination Chemistry Reviews**, 2016.

PASCHOAL, D. et al. Predicting Pt-195 NMR chemical shift using new relativistic all-electron basis set. **Journal of Computational Chemistry**, 2016.

SUN, Y. et al. Pt(II) and Au(III) complexes containing Schiff-base ligands: A promising source for antitumor treatment. **European Journal of Medicinal Chemistry**, 2021.

TANASE, Tomoaki; TANAKA, Miho; HAMADA, Mami; MORITA, Yuka; NAKAMAE, Kanako; URA, Yasuyuki; NAKAJIMA, Takayuki. Alloyed Tetranuclear Metal Chains of Pd_{4-n}Pt_n (n=0–3) Scaffolded by a New Linear Tetrakisphosphine Containing a PNP Bridge. **Chemistry – A European Journal**, [S.L.], v. 25, n. 35, p. 8219-8224, 15 maio 2019. Wiley.

TEHFE, M. A. et al. Cyclometallated Pt(II) complexes in visible-light photoredox catalysis: New polymerization initiating systems. **Macromolecular Chemistry and Physics**, v. 213, n. 21, 2012.

WEI, J. et al. Machine learning in materials science. **InfoMat**, 2019.