

Universidade de São Paulo
Instituto de Física de São Carlos

XII Semana Integrada do Instituto de
Física de São Carlos

Livro de Resumos

São Carlos
2022

Semana Integrada do Instituto de Física de São Carlos

SIFSC 12

Coordenadores

Prof. Dr. Osvaldo Novais de Oliveira Junior

Diretor do Instituto de Física de São Carlos – Universidade de São Paulo

Prof. Dr. Javier Alcides Ellena

Presidente da Comissão de Pós Graduação do Instituto de Física de São Carlos – Universidade de São Paulo

Profa. Dra. Tereza Cristina da Rocha Mendes

Presidente da Comissão de Graduação do Instituto de Física de São Carlos – Universidade de São Paulo

Comissão Organizadora

Adonai Hilario

Arthur Deponte Zutião

Elisa Goettems

Gabriel dos Santos Araujo Pinto

Henrique Castro Rodrigues

Jeffer Santiago Mares

João Victor Pimenta

Julia Martins Simão

Letícia Martinelli

Lorany Vitoria dos Santos Barbosa

Lucas Rafael Oliveira Santos Eugênio

Natasha Mezzacappo

Paulina Ferreira

Vinícius Pereira Pinto

Willian dos Santos Ribela

Normalização e revisão – SBI/IFSC

Ana Mara Marques da Cunha Prado

Maria Cristina Cavarette Dziabas

Maria Neusa de Aguiar Azevedo

Sabrina di Salvo Mastrantonio

Ficha catalográfica elaborada pelo Serviço de Informação do IFSC

Semana Integrada do Instituto de Física de São Carlos
(12: 10 out. - 14 out. : 2022: São Carlos, SP.)
Livro de resumos da XII Semana Integrada do Instituto de Física de São Carlos/ Organizado por Adonai Hilario [et al.]. São Carlos: IFSC, 2022.
446 p.
Texto em português.
1. Física. I. Hilario, Adonai, org. II. Título

ISBN: 978-65-993449-5-4 CDD: 530

PG160

Predição de interações fármaco-alvo com Biclustering Random Forests

ILIDIO, Pedro; ALVES, André Hallwas Ribeiro; THIEMANN, Otavio; CERRI, Ricardo

ilidio@alumni.usp.br

Diversos processos naturais e sociais se manifestam sob a forma de redes bipartidas, em que dois grupos de instâncias apresentam conexões apenas entre si, sem conexões internas a cada um. Exemplos de problemas passíveis de serem apresentados nesse formato incluem as mais diversas formas de interação entre dois tipos distintos de entidades, como interações fármaco-proteína, microRNA-mRNA, ou mesmo itens e usuários em um serviço de compras *on-line*. No domínio biológico, despertam grande interesse os processos de descoberta e reposicionamento de fármacos, em que novas interações entre moléculas pequenas e alvos no organismo (como proteínas receptoras) são determinadas. Contudo, a síntese dessas moléculas e os ensaios *in vitro* necessários para o aferimento de novas interações são de elevado custo, de maneira que novas metodologias *in silico* têm, cada vez mais, ganhado espaço na pesquisa científica. Dentre elas, abordagens de aprendizado de máquina (AM) têm se mostrado especialmente atrativas, com a grande vantagem de não necessitarem de simulações físicas computacionalmente custosas ou da resolução de estruturas proteicas, como o *docking* molecular. Entretanto, a grande quantidade de interações possíveis, bem como validação experimental frequentemente escassa e só capaz de confirmar interações positivas, dificultam enormemente sua aplicação. Além disso, alguns dos modelos mais profícuos do campo, que fazem uso de aprendizado profundo, são naturalmente incapazes de explicar as decisões tomadas, de forma que sua capacidade de resolver o problema de predição dificilmente se traduz em conhecimento sobre a tarefa, conhecimento esse que, disponível à humanidade, fomentaria o desenvolvimento metodológico inclusive fora do escopo do AM. Nesse contexto, trabalhos anteriores (1-2) provêm uma adaptação simples de métodos baseados em árvores de decisão, notadamente transparentes, para dados de interação, permitindo com que todas as interações candidatas possam ser consideradas em tempo plausível de construção do modelo. O presente trabalho estende essa linha de pesquisa ao introduzir otimizações e novas formas de ensemble, reimplementando as árvores adaptadas de forma eficiente e compatível com bibliotecas atuais, bem como apresentando as Biclustering Random Forests, versão do célebre algoritmo de Breiman (3) otimizada a dados bipartidos. Explora-se também a regularização do modelo com informações de vizinhança, buscando mitigar de forma inovadora o problema da esparsidade da matriz de adjacência ao introduzir influência de critérios não-supervisionados durante o crescimento das árvores em si, com mínima necessidade de pré-processamento dos dados. Nota-se que o teste e validação computacional dos modelos de AM no contexto de interações fármaco-alvo positivas-desconhecidas são frequentemente realizados de forma questionável, reportando pontuações incongruentes com a complexidade do problema em questão. São então providas e demonstradas ferramentas para auxiliar nesse processo, reavaliando e comparando métodos anteriores sob ótica mais criteriosa. Por fim, são obtidos resultados competitivos com os modelos estado-da-arte em quatro *datasets* padrão-ouro de *benchmarking*, e espera-se que as ferramentas aqui disponibilizadas possam facilitar o desenvolvimento de novas técnicas de predição de interações num futuro próximo.

Palavras-chave: Interações fármaco-alvo. Árvores de decisão semi-supervisionadas. Random Forests.

Agência de fomento: CAPES (88887.529627/2020-00; 88887.641930/2021-00; 88887.684441/2022-00)

Referências:

1. PLIAKOS, K.; GEURTS, P.; VENS, C. Global multi-output decision trees for interaction prediction. **Machine Learning**, v. 107, n. 8, p. 1257–1281, May 2018. Disponível em: <https://link.springer.com/article/10.1007/s10994-018-5700-x>. Acesso em: 9 set. 2022.
2. PLIAKOS, K.; VENS, C. Drug-target interaction prediction with tree-ensemble learning and output space reconstruction. **BMC Bioinformatics**, v. 21, n. 1, p. 1–11, 7 Feb.. 2020. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-3379-z>. Acesso em: 9 set. 2022.
3. BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001. Disponível em: <https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf>. Acesso em: 9 set. 2022.