Universidade de São Paulo
Instituto de Física de São Carlos

Semana Integrada do Instituto de Física de São Carlos

13ª edição

Livro de Resumos

São Carlos
2023

Ficha catalográfica elaborada pelo Serviço de Informação do IFSC

Semana Integrada do Instituto de Física de São Carlos
(13: 21-25 ago.: 2023: São Carlos, SP.)
Livro de resumos da XIII Semana Integrada do Instituto de Física de São Carlos – Universidade de São Paulo / Organizado por Adonai Hilário da Silva [et al.]. São Carlos: IFSC, 2023.
358p.

Texto em português.
1.Física. I. Silva, Adonai Hilário da, org. II. Título.

ISSN: 2965-7679

PG141

Florestas bipartidas semi-supervisionadas para predição de interações

CERRI, Ricardo[1]; ILÍDIO, Pedro[2]; ALVES, André Hallwas Ribeiro[1]; THIEMANN, Otavio Henrique[2]

ilidio@alumni.usp.br

[1]Universidade Federal de São Carlos - UFSCar; [2]Instituto de Física de São Carlos - USP

Diversas tarefas de aprendizado de máquina podem ser formuladas como a predição de interações entre pares de entidades, muitas vezes representando relações entre objetos de duas classes distintas. Exemplos de tais tarefas são abundantes na biologia molecular computacional e incluem a predição de interações entre proteínas, entre proteínas e fármacos, entre proteínas e RNAs longos não codificantes (lncRNA) e entre microRNAs e RNAs mensageiros. O presente trabalho visa estudar e aprimorar algoritmos de aprendizado especificamente voltados a resolver esse tipo de problema, propondo modificações que melhorem tanto sua performance de predição como sua complexidade computacional. Aspectos complicantes dos problemas de predição de interações são as altas dimensionalidade e esparsidade dos rótulos disponíveis, que se originam natureza quadrática do número de possíveis pares em relação ao número de entidades de cada tipo. Como consequência, uma parcela pequena das possiveis interações são experimentalmente verificadas e compõem o conjunto de dados de treinamento, e a maioria das interações são desconhecidas. Tal cenário, por vezes denotado positive-unlabeled learning, coloca nuâncias na forma como o treinamento dos modelos é realizado e sugere que abordagens semi-supervisionadas, em que agrupamentos de entidades compõem o processo de treinamento, podem apresentar vantagens conforme o número de interações desconhecidas aumenta. (1) Assim, propomos algoritmos baseados em árvores de decisão semi-supervisionadas que operam diretamente sobre redes bipartidas, e os comparamos com modelos já bem estabelecidos na literatura. Mostramos desempenho de predição competitivo com o estado-da-arte em diferentes tarefas de predição de interações, e ganhos no tempo de treinamento são demonstrados para as adaptações de algoritmo desenvolvidas em relação aos modelos originais. Espera-se que as ideias discutidas e ferramentas disponibilizadas possam fomentar o estudo das àrvores bipartidas e permitir que dados cada vez mais volumosos sejam levados integralmente em consideração.

Palavras-chave: Aprendizado de máquina. Predição de interações. Árvores de decisão.

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

Referências:

1 BEKKER, J.; DAVIS, J. Learning from positive and unlabeled data: a survey. **Machine Learning**, v. 109, n. 4, p. 719-760, Apr. 2020.