KNOWLEDGE-AUGMENTED GRAPH MACHINE LEARNING FOR DRUG DISCOVERY: FROM PRECISION TO INTERPRETABILITY

Zhiqiang Zhong & Davide Mottin

Aarhus University

OUTLINE

I. Introduction and Motivation
II. Background of Drug Discovery
III. Graph Machine Learning (GML) and Knowledge Graph (KG) in Drug Discovery
IV. Knowledge-augmented Graph Machine Learning (KaGML) for Drug Discovery
V. Practical Resources
VI. Open Challenges and Future Directions

OPEN CHALLENGES

1. Knowledge database composition and compatibility
   i. The effectiveness of KaGML methods heavily relies on the availability of qualified knowledge databases that can provide comprehensive and sufficient information.
   ii. Harmonisation and integration of data still pose a significant challenge, as these resources are often diverse, heterogeneous, and distributed across multiple platforms. As such, addressing the lack of standardisation in data integration is a critical area for future research to enhance the power of KaGML.
   iii. Many biomedical knowledge databases have to be frequently updated and refined to stay up to date with the current research, which presents a challenge for KaGML methods. To address this, it is recommended that KaGML works store the versions of the databases used in their experiments for better reproducibility.
   iv. Important principles: FAIR[1].

[1] *Mark D Wilkinson et al.*, The fair guiding principles for scientific data management and stewardship. Scientific Data, 2016

OPEN CHALLENGES

2. Effective knowledge integration with uncertainty
   i. KaGML works have incorporated external knowledge into *preprocessing*, *pretraining*, *training*, and *interpretability* for drug discovery.
   ii. However, these approaches are typically deterministic, ignoring the underlying uncertainty of knowledge and its impact on model learning and inference.
   iii. Thus, it is an important area of future research to investigate how to effectively and systematically model knowledge uncertainties for real-world applications.

OPEN CHALLENGES

3. Advanced interpretability & careful evaluation benchmark
   i. The enhancement of the interpretability of AI models has the potential to increase the confidence and dependability of patients, as well as to enhance the applicability of the models. Nevertheless, there remains a vast scope for further research in the area of advanced interpretability, with the aim of providing more holistic and adaptable explanations, such as advanced reasoning and question-answering capabilities.
   ii. Designing a comprehensive validation pipeline, such as an explanation verification pipeline, is a promising area for future research. While KaGML approaches have been developed to address interpretability problems in drug discovery, the question of how to verify and evaluate the generated explanations remains open.

OPEN CHALLENGES

4. From drug discovery to more biomedical fields
   i. While this tutorial focuses on the recent advancements in drug discovery, other fields of biomedical research could benefit from the expanding use of KaGML techniques, including target identification and validation and gene and cell therapy.
   ii. It would be interesting to see the development of a unified KaGML framework that supports diverse healthcare services.

OPEN CHALLENGES

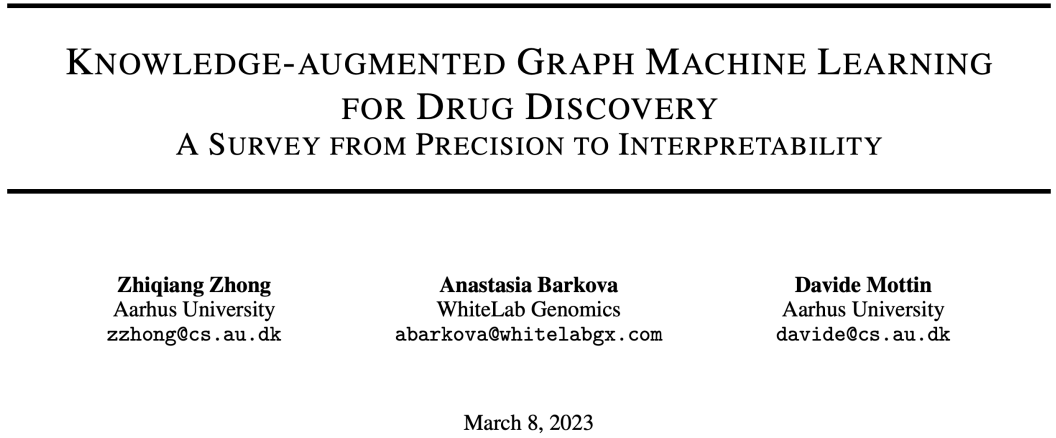
5. Security & privacy and efficiency & scalability
   i. The advancements in machine learning and growth in computational capacities have transformed the technology landscape but have also raised concerns about security and privacy.
   ii. This includes guaranteeing the ownership of knowledge databases, protecting patient-sensitive information, and ensuring the viability of models against malicious attacks.

OPEN CHALLENGES

1. Knowledge database composition and compatibility
2. Effective knowledge integration with uncertainty
3. Advanced interpretability & careful evaluation benchmark
4. From drug discovery to more biomedical fields
5. Security & privacy and efficiency & scalability

NEXT?

- A related survey manuscript is available online at:
  https://arxiv.org/abs/2302.08261



KNOWLEDGE-AUGMENTED GRAPH MACHINE LEARNING FOR DRUG DISCOVERY
A SURVEY FROM PRECISION TO INTERPRETABILITY

Zhiqiang Zhong
Aarhus University
zzhong@cs.au.dk

Anastasia Barkova
WhiteLab Genomics
abarkova@whitelabgx.com

Davide Mottin
Aarhus University
davide@cs.au.dk

March 8, 2023

- Detailed information about KaGML survey/papers/practical resources:
  https://github.com/zhiqiangzhongddu/Awesome-Knowledge-augmented-GML-for-Drug-Discovery

ACKNOWLEDGEMENT

Zhiqiang Zhong
Postdoc
Aarhus University

Davide Mottin
Asst. Prof.
Aarhus University

Our work is supported by the Horizon Europe and Danmarks Innovationsfond under the Eureka, Eurostar grant no E115712.

Thank you!

Questions?

zzhong@cs.au.dk
https://zhiqiangzhongddu.github.io/

AARHUS
UNIVERSITY