



LICENCIATURA EN CIENCIAS DE DATOS
Propuesta de tema de tesis

Redes convolucionales en grafos con aplicación al problema de reposicionamiento de fármacos

El presente proyecto planea estudiar metodologías que permitan extraer información de relevancia biomédica a partir de la integración de datos heterogéneos. Así, utilizaremos el abordaje de redes convolucionales montadas sobre grafos (Graph Convolutional Networks) para predecir nuevas posibles relaciones entre genes, enfermedades de origen genético, dianas terapéuticas y fármacos. Utilizando redes quimio-genómicas, es posible recomendar nuevos targets terapéuticos para drogas conocidas y ya aprobadas? Utilizando redes de asociaciones gen-enfermedad, es posible inferir nuevas asociaciones entre genes y enfermedades de base-hereditaria? Estas son algunas de las preguntas que nos interesan responder.

Palabras clave: grafos de conocimiento, redes convolucionales en grafos, reposicionamiento de farmacos

Conocimientos deseables

Conocimientos de python, libreria pytorch, conocimientos de R

¿Qué podría aprender quien realice esta tesis?

Manejo de redes complejas para integración de datos, redes convolucionales, tareas de priorizacion en redes

Dirección de la tesis

*Chernomoretz, Ariel
Dto Fisica, FCEN, UBA*

Contacto: ariel@df.uba.ar

Más información en el pdf a continuación.

Redes convolucionales en grafos con aplicación al problema de reposicionamiento de fármacos

INTRODUCCIÓN

El interés de la industria bio-farmacéutica en enfoques computacionales para el desarrollo de nuevos productos farmacológicos está motivada por los altísimos costos vinculados con llevar una nueva droga al mercado [Scannell2012, Pushpakom2019]. La alternativa encarnada en el enfoque de reposicionamiento de drogas contempla la identificación *in-silico* de nuevos usos para drogas ya aprobadas. Este abordaje supone enormes ventajas en términos de abaratamiento de costos, reducción de riesgos y aceleración de tiempos de desarrollo [Parvathaneni2019]. Las estrategias implementadas para este tipo de tareas involucran muchas veces técnicas de aprendizaje automático, difusión en redes y aprendizaje profundo para relacionar enfermedades, genes, fenotipos, drogas y estructuras moleculares, a la hora de reposicionar fármacos [Park2019, Issa2021, Sadeghi2021].

Para atacar problemáticas como la descrita arriba un abordaje basado en redes resulta particularmente atractivo. Dicho enfoque ofrece una representación natural de la intrincada red de asociaciones de relevancia biológica que pueden ser utilizadas para condensar información sobre similitud o relaciones establecidas entre pares de entidades relevantes. Por ejemplo: genes que regulan a otros genes, proteínas que interactúan formando complejos proteicos, drogas de estructura química similar, enfermedades que comparten síntomas, etc. Al mismo tiempo, las redes pueden utilizarse para integrar conocimiento de diferentes dominios vinculando, mediante relaciones específicas, a nodos de diferente naturaleza (por ejemplo: genes con enfermedades de base hereditaria, proteínas con funciones biológicas o drogas con targets proteicos). Conocidos como grafos-de-conocimiento (knowledge-graphs), se presentan de esta manera como herramientas muy eficientes para integrar datos heterogéneos en una estructura formal que permita utilizar de manera efectiva la información embebida para producir predicciones sobre nuevas asociaciones o nuevas hipótesis biológicas testeables [Sadegh2021, FernandezTorras2022, Chandak2023].

La estrategia de priorización que nos interesa considerar tiene que ver con una variante de metodologías de aprendizaje automático profundo, acopladas a grafos de conocimiento llamada Graph Convolutional Networks (GCN). En efecto, en los últimos años se han desarrollado métodos de aprendizaje automático que extienden el concepto de redes neuronales convolucionales a su aplicación en grafos [Kipf2016]. Estas técnicas lograron una performance comparable o superior al estado del arte en tareas de aprendizaje semi-supervisado sobre redes complejas, lo que motivó el desarrollo de múltiples arquitecturas y modelos que se adaptan a bases de datos de gran escala [Hamilton2017], son generalizables a redes heterogéneas [Wang2019, Schlichtkrull2018] y permiten incorporar mecanismos atencionales [Veličković2018]. Debido a su capacidad de capturar relaciones complejas entre objetos de naturaleza heterogénea, existe un interés creciente en la aplicación de redes convolucionales a bases de datos biomédicas [Zitnik2018, Li2019]. Creemos que el uso de este tipo de abordajes combina la capacidad descriptiva y el poder de las redes complejas para integrar capas de conocimiento heterogéneo, con un potente motor de inferencia provisto por técnicas de aprendizaje profundo, por lo que es nuestro interés avanzar dentro de este marco para implementar las tareas de reposicionamiento y priorización planteadas.

METODOLOGIA

El trabajo consistirá en implementar una GCN en un grafo de conocimiento que relacione drogas y moléculas pequeñas con proteínas. Para ello

1. se realizará una breve inmersión bibliográfica que incluirá una rápida introducción a los conceptos biomédicos relevantes y a la teoría detrás de las GCN
2. se construirá primeramente un grafo de conocimiento con datos obtenidos de repositorios públicos (ChEMBL, DrugBank, PubChem) donde se vinculen compuestos químicos con targets proteicos a partir de bioactividades reportadas
3. A partir de una librería desarrollada en el grupo (basada en PyTorch) se implementará una GCN para sugerir nuevos posibles targets terapéuticos a drogas ya conocidas.

RESULTADOS ESPERADOS

Al final del trabajo se pretende contar con una implementación funcional de una GCN que permita realizar las priorizaciones requeridas.

DEDICACION

Se prevé una dedicación de 20 hs semanales, con participación en las reuniones de grupo

REFERENCIAS

[Bronstein2021] Bronstein et al, *Geometric deep learning: Grids, groups, graphs, geodesics and gauges*, arXiv:2104.13478

[Chandak2023] Chandak, P., Huang, K. & Zitnik, M. Building a knowledge graph to enable precision medicine. *Sci Data* 10, 67 (2023)

[FernándezTorras2022] Fernández-Torras, A., Duran-Frigola, M., Bertoni, M. et al. Integrating and formatting biomedical data as pre-calculated knowledge graph embeddings in the Bioteque. *Nat Commun* 13, 5304 (2022)

[Hamilton2017] Inductive Representation Learning on Large Graphs, Hamilton W, Ying R, Leskovec J. (2017). In *Advances in Neural Information Processing Systems* 30,

[Issa2021] Issa NT, Stathias V, Schürer S, Dakshanamurthy S. Machine and deep learning approaches for cancer drug repurposing. *Semin Cancer Biol.* 2021 Jan;68:132-142

[Jia2021] Jia J, Benson A., A Unifying Generative Model for Graph Learning Algorithms: Label Propagation, Graph Convolutions, and Combinations. arXiv:2101.07730, 2021

[Kipf2016] Semi-Supervised Classification with Graph Convolutional Networks, Kipf T and Welling M (2016), In *International Conference on Learning Representations* (2016),

- [Li2019] PGCN: Disease gene prioritization by disease and gene embedding through graph convolutional neural networks. Li, Y., Kuwahara, H., Yang, P., Song, L., & Gao, X. (2019), bioRxiv,<https://doi.org/10.1101/532226>
- [Parvathaneni2019] Parvathaneni V, Kulkarni NS, Muth A, Gupta V. Drug repurposing: a promising tool to accelerate the drug discovery process. *Drug Discov Today*. 2019 Oct;24(10):2076-2085.
- [Park2019] Park, K. A review of computational drug repurposing. *Transl. Clin. Pharm.* 27, 59–63 (2019).
- [Pushpakom2019] Pushpakom, S. et al. Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18, 41–58 (2019).
- [Sadeghi2021] Sadeghi SS, Keyvanpour MR. An Analytical Review of Computational Drug Repurposing. *IEEE/ACM Trans Comput Biol Bioinform.* 2021 Mar-Apr;18(2):472-488.
- [Scannell2012] Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* 11, 191–200 (2012).
- [Schlichtkrull2018] Modeling Relational Data with Graph Convolutional Networks, Schlichtkrull M, Kipf T, Bloem P, van den Berg R, Titov I, Welling M. (2018). In: , et al. *The Semantic Web. ESWC 2018. Lecture Notes in Computer Science()*, vol 10843. Springer, Cham. https://doi.org/10.1007/978-3-319-93417-4_38
- [Veličković2018] Graph Attention Networks, Veličković P, Cucurull G, Casanova A, Romero A, Li Petro and Bengio Y. In *International Conference on Learning Representations* (2018), <https://doi.org/10.48550/arXiv.1710.10903>
- [Wang2019] Heterogeneous Graph Attention Network, Wang X, Ji H, Shi C, Wang B, Ye Y, Cui Pi, and Yu P (2019). In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 2022–2032.
- [Wang2021] Wang H, Leskovec J, Combining Graph Convolutional Neural Networks and Label Propagation, *ACM Transactions on Information Systems*, 40-4 (2021)
- [Zitnik2018] Modeling polypharmacy side effects with graph convolutional networks, Zitnik M, Leskovec J, Agrawal M (2018), *Bioinformatics*, Volume 34, Issue 13, 01 July 2018, Pages i457–i466,