



A Review on Disease Prediction System

Tejas Gorakh Mahale
MSC (Computer Science)
Indira College, Malegaon

Prof. P.V. Shewale
MSC (Computer Science)
Indira College, Malegaon

Abstract-: Increasing amounts of evidence indicate that lncRNAs are implicated in various complex human diseases. The small fraction of experimentally confirmed lncRNA-disease associations has created an urgent need for computational prediction models. Although numerous approaches have been proposed, there remains significant room for improvement. To tackle the cold-start problem and accurately represent associations, this paper treats the prediction of lncRNA-disease associations as a recommendation problem and introduces a method based on matrix factorization and neural networks. First, to better represent lncRNAs and diseases, their embeddings are learned through matrix factorization. Then, features of associations are represented by integrating embeddings of lncRNAs and diseases. Finally, a neural network is used to predict potential associations.

Keywords-: Disease associations, Computational prediction model, Matrix factorization, Neural network

I. INTRODUCTION

In the modern world, the importance of proactive healthcare cannot be exaggerated. Early detection and diagnosis of diseases can significantly improve patient outcomes and reduce healthcare costs. This project aims to harness the power of Python and Machine Learning to predict diseases based on various health parameters. We leverage Python, a versatile programming language popular in the data science community, due to its simplicity and the availability of robust libraries such as Pandas for data manipulation, Scikit-learn for machine learning. Our machine learning models are trained on a comprehensive dataset, which includes a wide range of health parameters. These models are capable of identifying patterns and learning from them, thereby predicting the likelihood of a disease. The goal of this project is not only to build a predictive model but also to interpret the model's predictions, which can provide valuable insights into disease risk factors. This could potentially lead to better prevention strategies and more effective treatments.

II. LITERATURE SURVEY

Gupta et al [1], this paper has designed a smart trolley with the help of Arduino and also implemented the feature of security in it. The outline of the trolley designed was like a mailbox when an item is dropped in it, the entry used to get shuttered. Automatically the entry will open only when the payment of the items purchased was paid, but this paper had many disadvantages such as once the product is added into the trolley and if the customer does not want to buy it, then the doors will not open unless payment is done.

Kiran Dhokale et al [2], this paper has proposed to implement SQLite as a database server for smartphones android applications. The SQLite database is not a centralized database, this tends to be a drawback for the paper. so the datasets managed in the application's database will be local to the device.

Ansar Ahmad et al [3], this paper explains the significance of Location-Based structure which will set aside our time during multiple tasks like finding specific addresses of some stores, getting and knowing about some good offers on the different items at a time. By using this mobile android application, customers can able to search for different Shops, Stores, Groceries locations.

Prakruthi K et al [4], In this paper, RFID tags have been used instead of barcodes. The RFID tag will be applied to the items in the stores or supermarkets. Whenever the customer drops an item into the trolley, it will be inspected and examined by the RFID reader and item details such as name and price will be displayed on the visual display screen.

portion of the lexicon is dropped, and a fingerprint is generated by hashing the document once for each variant of the lexicon. This results in each document being represented by a vector of hash values. When testing two documents for near-duplicate status, if any of the matching pairs of elements of their vectors collide, the documents can be considered near-duplicates.

Henzinger (2021) proposed detecting near-duplicates by combining shingling and locality sensitive hashing. The method begins by using shingling to identify near-duplicate candidates, and then filters these results using locality sensitive hashing to identify near-duplicate documents. The author obtained promising results on a very large dataset, both for near-duplicate documents on the same website and near-duplicate documents on different websites.

III. SYSTEM ARCHITECTURE

The entire process of our proposed approach is summarized in Fig. 1 Flowchart of the research pipeline. The miRNA-lncRNA interactions and miRNA-disease associations are exploited for the construction of the tripartite graph. The tripartite graph, in its turn, is at the basis of both neighborhood analysis and collaborative filtering steps, from which the three proposed approaches are obtained: NGH

from neighborhood analysis, CF from collaborative filtering, NGH-CF from the combination of the two ones. Each prediction approach returns in output a LDAs rank

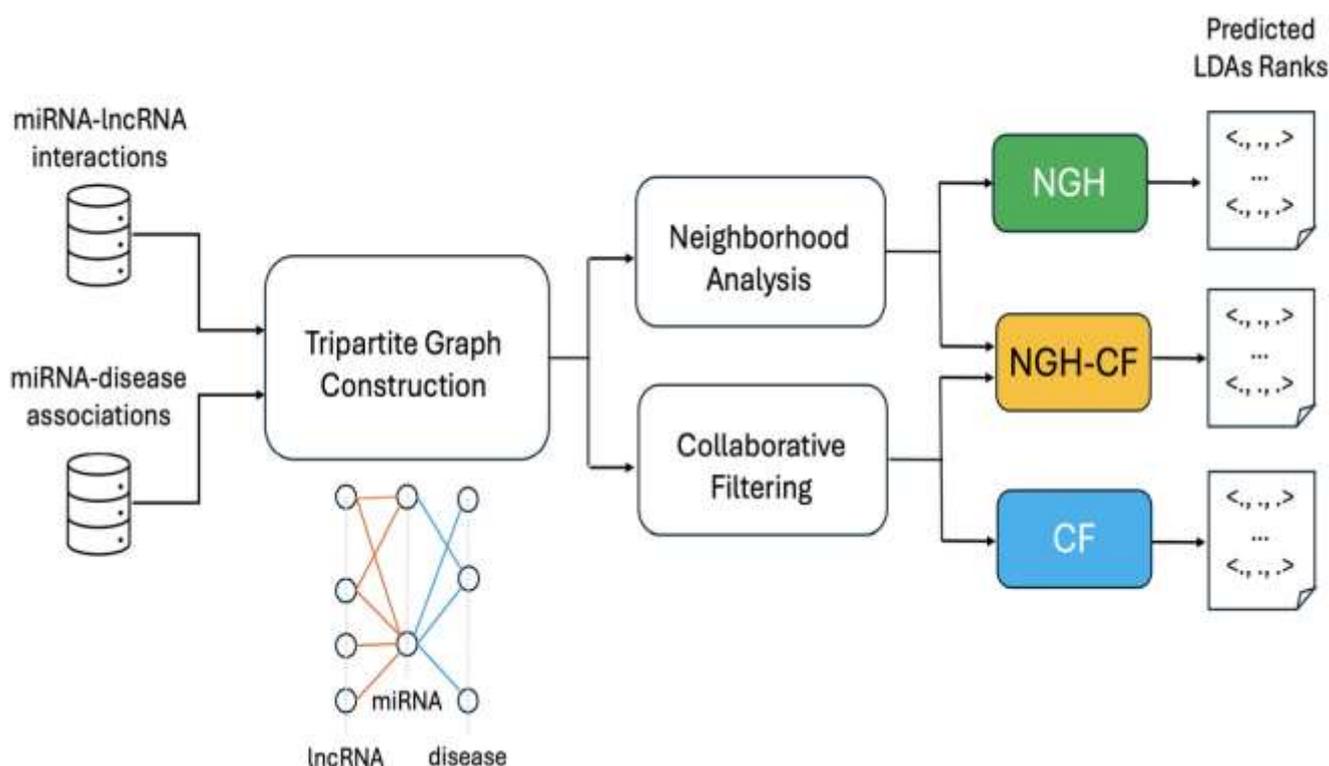


Figure 1. Detail System architecture

The proposed approaches have been thoroughly validated on both synthetic and real datasets, demonstrating that they significantly outperform other methods from the literature. The experimental analysis indicates that the improved accuracy of the proposed methods stems from their ability to capture specific scenarios overlooked by competitors. For instance, our approaches can detect true LDAs that other methods miss, where the involved lncRNA does not share intermediate molecules with the associated disease, despite neighboring lncRNAs sharing numerous miRNAs with that disease. Additionally, our methods are shown to be robust to noise introduced by perturbing a controlled percentage of lncRNA-miRNA interactions and miRNA-disease associations, with NGH-CF being the most robust. The experimental results suggest that the proposed prediction methods can effectively aid biologists in selecting significant associations for further laboratory verification.

CONCLUSION

Accumulating evidence suggests that lncRNAs play crucial roles in many complex human diseases. Identifying potential lncRNA-disease associations can enhance our understanding of the pathogenesis of these diseases. Consequently, various computational approaches have been proposed to predict associations between lncRNAs and diseases. To address the cold-start problem and accurately represent associations, we present a method based on matrix factorization and neural networks. Matrix factorization is used to learn embeddings of lncRNAs and diseases, and features of associations are constructed from these embeddings. A neural network is then applied to the prediction task, using these association features as inputs. This approach is compared with state-of-the-art methods.

REFERENCE

1. T. R. Mercer, M. E. Dinger, and J. S. Mattick, "Long non-coding RNAs: insights into functions," *Nature Reviews Genetics*, vol. 10, no. 3, pp. 155–159, 2009.
2. K. C. Wang and H. Y. Chang, "Molecular mechanisms of long noncoding RNAs," *Molecular Cell*, vol. 43, no. 6, pp. 904–914, 2011.
3. O. Wapinski and H. Y. Chang, "Long noncoding RNAs and human disease," *Trends in Cell Biology*, vol. 21, no. 6, pp. 354–361, 2011.
4. L. W. Harries, "Long non-coding RNAs and human disease," *Biochemical Society Transactions*, vol. 40, no. 4, pp. 902–906, 2012.
5. N. Ishii, K. Ozaki, H. Sato, H. Mizuno, S. Saito, A. Takahashi, Y. Miyamoto, S. Ikegawa, N. Kamatani, M. Hori, S. Saito, Y. Nakamura, and T. Tanaka, "Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction," *Journal of Human Genetics*, vol. 51, no. 12, pp. 1087–1099, 2006.
7. R. Johnson, "Long non-coding RNAs in Huntington's disease neurodegeneration," *Neurobiology of Disease*, vol. 46, no. 2, pp. 245–254, 2012.
8. S. U. Schmitz, P. Grote, and B. G. Herrmann, "Mechanisms of long noncoding RNA function in development and disease," *Cellular and Molecular Life Sciences*, vol. 73, no. 13, pp. 2491–2509, 2016.

9. J. Li, Z. Xuan, and C. Liu, "Long non-coding RNAs and complex human diseases," *International Journal of Molecular Sciences*, vol. 14, no. 9, pp. 18 790–18 808, 2013.
10. L. Ding, M. Wang, D. Sun, and A. Li, "TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph," *Scientific Reports*, vol. 8, pp. 1–11, 2018.
11. P. Ping, L. Wang, L. Kuang, S. Ye, M. F. B. Iqbal, and
12. T. Pei, "A novel method for lncRNA-disease association prediction based on an lncRNA-disease association network," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 2, pp. 688–693, 2019.
13. C. Gu, B. Liao, X. Li, L. Cai, Z. Li, K. Li, and J. Yang, "Global network random walk for predicting potential human lncRNA-disease associations," *Scientific Reports*, vol. 7, no. 1, pp. 1–11, 2017.
14. Y. Gao, J. Hu, and X. Shang, "Identification of lncRNA-disease association using bi-random walks," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 1249–1255
15. J. Li, H. Zhao, Z. Xuan, J. Yu, X. Feng, B. Liao, and L. Wang, "A novel approach for potential human lncRNA-disease association prediction based on local random walk," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019.
16. X. Chen and G. Yan, "Novel human lncRNA-disease association inference based on lncRNA expression profiles," *Bioinformatics*, vol. 29, no. 20, pp. 2617–2624, 2013.
17. C. Lu, M. Yang, F. Luo, F. X. Wu, M. Li, Y. Pan, Y. Li, and J. Wang, "Prediction of lncRNA-disease associations based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 19, pp. 3357–3364, 2018.
18. M. Zeng, C. Lu, F. Zhang, Z. Lu, and M. Li, "lncRNA-disease association prediction through combining linear and non-linear features with matrix factorization and deep learning techniques," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019, pp. 577–582.
19. M. Zeng, C. Lu, Z. Fei, F. Wu, Y. Li, W. Jianxin, and L. Min, "DMFLDA: A deep learning framework for predicting lncRNA-disease associations," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022.
20. G. Chen, Z. Wang, D. Wang, C. Qiu, M. Liu, X. Chen,
21. Q. Zhang, G. Yan, and Q. Cui, "lncRNADisease: a database for long-non-coding RNA-associated diseases," *Nucleic Acids Research*, vol. 41, no. D1, pp. D983–D986, 2013.
22. Z. Lu, K. B. Cohen, and L. Hunter, "GeneRIF quality assurance as summary revision," *Pacific Symposium on Biocomputing*, pp. 269–280, 2007.
23. S. Ning, J. Zhang, P. Wang, H. Zhi, J. Wang, Y. Liu,
24. Y. Gao, M. Guo, M. Yue, L. Wang, and X. Li, "lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers," *Nucleic Acids Research*, vol. 44, no. D1, pp. D980–D985, 2016.
25. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen,
26. C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin,
27. S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, and et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2015.
28. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.
29. P. Xuan, K. Han, Y. Guo, J. Li, X. Li, Y. Zhong,
30. Z. Zhang, and J. Ding, "Prediction of potential disease-associated microRNAs based on random walk," *Bioinformatics*, vol. 31, no. 11, pp. 1805–1815, 2015.
31. D. Sun, A. Li, H. Feng, and M. Wang, "NTSMDA: prediction of miRNA–disease associations by integrating network topological similarity," *Molecular Biosystems*, vol. 12, no. 7, pp. 2224–2232, 2016.
32. R. J. Grissom and J. J. Kim, *Effect Sizes for Research: A Broad Practical Approach*. Lawrence Erlbaum Associates, 2005.