Utilizing Graph Embeddings for Multiple Sclerosis Disease Modifying Therapy Adverse Event Prediction

Jason Patterson[1]

[1] **Tatonetti Lab, Department of Biomedical Informatics, Columbia University**

Background

Multiple sclerosis (MS) is one of the leading causes of disability among young adults. Currently, the administration of disease-modifying therapy (DMT) through immunosuppressive drugs is the front-line management strategy for the disease. However, the use of higher efficacy DMT frequently leads to the occurrence of adverse drug events (ADEs), and both risks and benefits must be assessed when selecting a treatment plan[1]. ADE information predominantly originates from clinical trials and post market surveillance via the Federal Adverse Events Reporting Service (FAERS)[2]. However, both methods are limited to population-level statistics and fail to address ADE risk on the individual level. In this work, we use sequences of OMOP medical concepts in electronic health records (EHRs) to create an individual-level predictive model for ADE occurrence related to MS DMT. With the aid of Graph Convolutional Networks (GCN), we also created a novel method KG-LIME for explaining the output of the predictive model in terms of feature importance. We attempted to address gaps in other EHR ADE prediction methods, including inattention to time-varying nature of EHR data, stringent feature selection methods, and failure to address causation.

Methods

First, we sought to train embedding representations of OMOP concepts relevant to the task of ADE prediction with GCNs, with the hope that these embeddings would have high interpretability. GCNs learn embeddings of nodes in a knowledge graph by sharing weights along the edges of the graph. The Relational Graph Convolutional Network (R-GCN) architecture and the Hierarchy-Aware Knowledge Graph (HAKE) method were selected to perform the training due to their high representative ability of node relationships[3,4]. As input, we created a ADE knowledge graph from a combination of four data sources: the OnSIDES ADE database, the Drug Central database , the LOINC2HPO project, and the OMOP CDM[5,6,7]. Overall, there were 12,676 nodes representing OMOP Condition, Drug, and Measurement standard concepts with 186,594 edges of the following types: *has_adverse_effect*, *has_indication*, *is_a*, *has_atc*, *associated_with_high* (measurement), and *associated_with_low* (measurement). A subgraph centered around the drug insulin is shown in Figure 1. The R-GCN was trained via negative sampling link prediction such that embeddings could differentiate negative false edges from true positive edges at a high rate.

We then designed a Long Short Term Memory (LSTM) predictive model for ADE occurrence that would take in as input the sequence of OMOP Condition, Drug, and Measurement concepts that are present in the knowledge graph and precede an index MS DMT drug administration date. Each concept was represented by its trained graph embedding and a “days removed from index” feature. The learning objective was the occurrence of an adverse event within 365 days following the index MS DMT drug administration date. A linear calibration layer at the end of the LSTM model ensured that outputs would be calibrated probabilities of ADE occurrence. We trained a separate LSTM model for 56 different adverse event types.

Finally, we created a novel model explanation method that is derived from the popular Local Interpretable Model-Agnostic Explanations (LIME) method[8]. This new method KG-LIME outputs influential concepts for ADE risk prediction by probabilistically replacing concepts in input sequences with concepts that have similar or distal embeddings (Figure 2). Embedding similarity is calculable due to the nature of the GCN training and can be determined in terms of the relationship type (e.g. concepts can be similar because they have similar *has_indication* edges or instead because they have similar *is_a* edges). The perturbed sequences are then used to re-predict ADE risk. A Decision Tree Regressor learns the re-predicted ADE risks using the numerical differences in perturbed embeddings as input. Gini feature importance is extracted from the Decision Tree and the top concepts are presented as an explanation of how perturbing the input to varying degrees can shift the decision boundary.

Results

For a set of 4,859 patients, we found that our model was effective at predicting 35 out of 56 adverse event types ($p < 0.05$) when compared to demographics and past diagnosis as variables. We also assessed discrimination and calibration in the form of the Area Under the Precision-Recall Curve (AUC-PR = 0.35 ± 0.25) and Brier Score (0.03 ± 0.02), respectively. These metrics seemed to vary across adverse event types and seemed highly correlated with frequency of event.

Additionally, KG-LIME generated relevant concepts for ADE prediction (Figure 3). Particularly, past medical history of an ADE and drugs that are either indicated or known to have adverse effects often appear in explanations. We performed a soft validation by confirming that highlighted concepts generally compared to observations cited in literature. We hope to perform a more formal validation in the future.

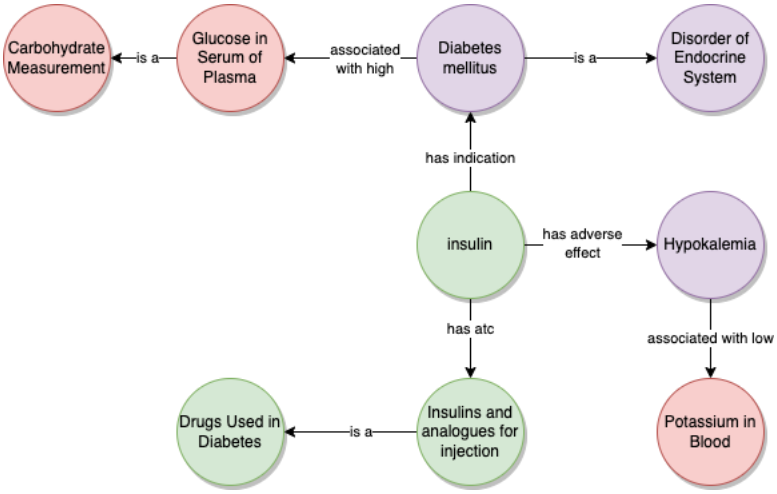

Figure 1: Insulin Subgraph of Knowledge Graph used to Train Concept Embeddings

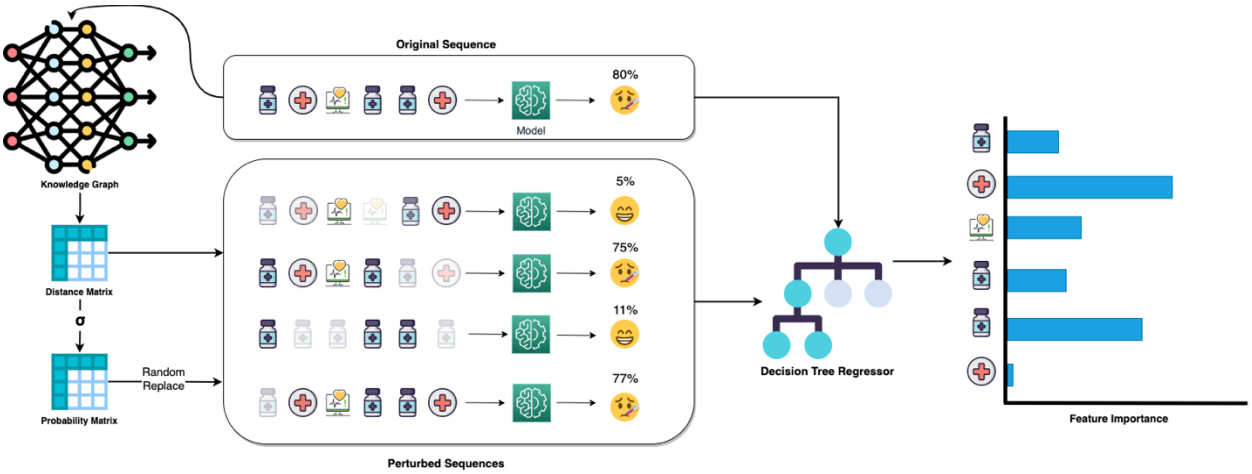

Figure 2: KG-LIME Model Explanation Algorithm

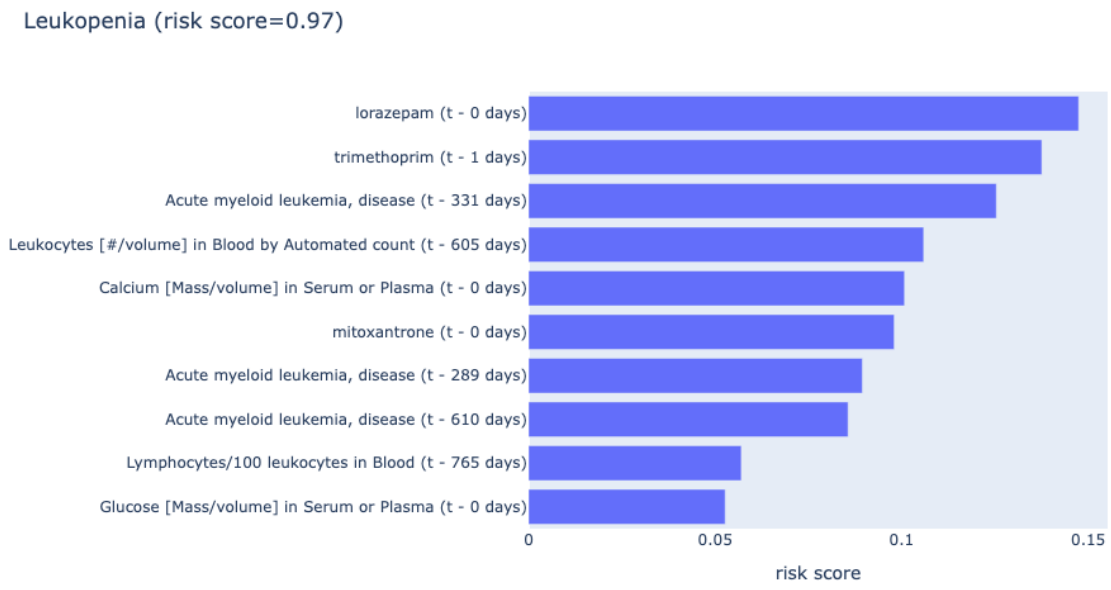

Figure 3: KG-LIME Explanation for Model Prediction of Leukopenia for a Single Patient

Conclusion

Many of our risk models demonstrated high calibration and discrimination for adverse event prediction. Furthermore, our novel KG-LIME method was able to utilize a knowledge graph to highlight OMOP concepts that were important to prediction. Future work will be required to further explore the temporal window of adverse event occurrence beyond the generic 1-year window used here, particularly for short-term inpatient adverse events and long-term severe adverse events. Additionally, we will need a more formal evaluation method for KG-LIME explanations.

References

1. He A, Merkel B. Timing of high-efficacy therapy for multiple sclerosis: a retrospective observational cohort study. The Lancet Neurology, vol. 19, pp. 307–316, Apr. 2020.
2. FDA Adverse Event Reporting System (FAERS) Public Dashboard. FDA. Oct. 2021. Publisher: FDA. https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard
3. M. Schlichtkrull, T. N. Kipf, P. Bloem, R. v. d. Berg, I. Titov, and M. Welling, “Modeling Relational Data with Graph Convolutional Networks,” Oct. 2017. arXiv:1703.06103 [cs, stat].
4. Z. Zhang, J. Cai, Y. Zhang, and J. Wang, “Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction,”Apr. 2022. arXiv:1911.09419 [cs, stat].
5. N. Tatonetti, “tatonetti-lab/onsides: OnSIDES Data Release v2.0.0-20230309,” Mar. 2023. https://github.com/tatonetti-lab/onsides/
6. “DrugCentral: online drug compendium | Nucleic Acids Research | Oxford Academic.” https://drugcentral.org/
7. X. A. Zhang, A. Yates, N. Vasilevsky, J. P. Gourdine, T. J. Callahan, L. C. Carmody, D. Danis, M. P. Joachimiak, V. Ravanmehr, E. R. Pfaff, J. Champion, K. Robasky, H. Xu, K. Fecho, N. A. Walton, R. L. Zhu, J. Ramsdill, C. J. Mungall, S. KÅNohler, M. A. Haendel, C. J. McDonald, D. J. Vreeman, D. B. Peden, T. D. Bennett, J. A. Feinstein, B. Martin, A. L. Stefanski, L. E. Hunter, C. G. Chute, and P. N. Robinson, “Semantic integration of clinical laboratory tests from electronic health records for deep phenotyping and biomarker discovery,” NPJ digital medicine, vol. 2, p. 32, 2019.
8. M. T. Ribeiro, S. Singh, and C. Guestrin, “”Why Should I Trust You?”: Explaining the Predictions of Any Classifier,” Aug. 2016. arXiv:1602.04938 [cs, stat].