Granularity-enhanced Ontology Rollup on Structured EHR Data to Mitigate Bias in Prediction Models

Jifan Gao, MS[1], Guanhua Chen, PhD[1]
[1]University of Wisconsin-Madison, Department of Biostatistics and Medical Informatics

Abstract: We developed a granularity-enhanced ontology rollup pipeline to mitigate performance bias for machine learning models that are developed using structured EHR data. Ontology rollup is a powerful method to encode structured EHR data. However, important details can be lost as all the medical concepts are rolled up to their parent concepts. Choosing the proper level of granularity of the features is important for risk prediction as the same features for different groups of the population may not be measured at the same precision. Our method adaptively increases granularity on features that are essential to a specific subgroup in order to provide more necessary details on this subgroup to a machine learning model. As the granularity is enhanced before the model training process, the method is model-agnostic and task-agnostic. Our experiment on the MIMIC-III dataset demonstrates that this method can improve discriminative power on the black group for a mortality prediction task.

Link to Video: https://vimeo.com/803724189
Link to Github page: https://github.com/GGGGFan/Granularity-Enhanced-Ontology-Rollup

Methodology Overview

In recent years, there has been a growing interest in using structured Electronic Health Records (EHR) data for clinical research informatics. Structured EHR data consists of clinical observations from various domains represented using standardized medical terms and stored in related tables. The use of standardized vocabularies enables the collection and organization of various data resources (medical conditions, drug exposures, lab measurements, etc.) in a common format, facilitating the systematic analysis of clinical observations. Machine learning technology has been widely used for clinical risk prediction using EHR data [1,2,3], but one of the challenges is the high dimensionality of the data and the existence of synonyms in standardized vocabularies [4].

The ontology rollup method is a powerful approach to handling numerous distinct medical concepts in EHR data. This method utilizes the parent-child relations in standardized vocabularies to aggregate synonymous concepts to their parent concept, reducing the dimensionality of the data set. For example, all types of diabetes (type I/II, with/without complications) from different vocabularies (ICD-9, ICD-10, SNOMED, etc.) will be converted to and represented as one ICD-9 section: diabetes mellitus. Machine learning pipeline integrated with this technology has been proven to perform well for risk prediction tasks and won several prestigious EHR-based data competitions in recent years [5,6]. Although effective, important details can sometimes be lost through the rollup process, potentially undermining the performance of a predictive model. For example, diabetes with complications leads to higher risks of mortality compared to diabetes without complications [7] but the occurrence of complications is ignored by the ontology rollup process. On the other hand, tracing all the aggregated concepts down to a more granular level will again lead to high dimensionality and hinder the generalizability of the model. Therefore, in order to provide more informative features to a model without overly increasing the model size, we need to adaptively construct features from structured EHR dataset with the right granularity.

In our proposed method, we aim to mitigate the performance disparity of ontology rollup-based machine learning models by increasing granularity on the underperformed group adaptively. Given a dataset and a trained machine learning model, our goal is to improve the model's performance on a subgroup that is underperforming compared to other subgroups, typically the minority group. The machine learning model is constructed using the ontology rollup technology. We first select the top N features that contribute the most to the minority group, where N is a hyperparameter that can be tuned during model development. Then, using the standard vocabulary, we trace down the aggregated medical conditions to collect their child concepts. These child concepts are added to the feature set, and a new machine-learning model is constructed using the new feature set. In particular, we "zoom in" on the features that contribute more to the predicted results. In the previous example, if the ICD-9 section of diabetes mellitus is a top contributing feature, we add all the subconcepts of diabetes mellitus to the feature set, including diabetes mellitus without mention of complication, diabetes with ketoacidosis, diabetes with hyperosmolarity, etc. Through the adaptive granularity enhancement process, we expect to provide details that are important for the minority group specifically so that the model performance can be elevated for this group. We validate our method on the MIMIC-III datasets and found that the AUPR and AUROC of the black group, on which a prediction model underperforms, increased from 0.137 and 0.810 to 0.144 and 0.816 on a mortality prediction task after applying our enhanced-granularity method.

Value Proposition

Our work is a novel procedure which leavage an newly proposed pipeline called ontology rollup for mitigating bias in risk prediction models using EHR data. The ontology rollup pipeline has been proven to be effective in several data challenge as a dimension reduction tools for predicting various phenotype. The proposed granularity-enhanced ontology rollup aims at solving the loss-of-detail issue on a specific subgroup and mitigating performance bias when implementing the original ontology rollup pipeline in clinical practice. Subconcepts of highly predictive concepts are added to the feature set for final prediction model building such that it could take the advantage of a more detailed feature representation for a given (minority group) sample. Our experiment on the MIMIC-III dataset demonstrates that the model's discrimination power (in particular, AUPR and AUROC) on a specific subgroup can be elevated with the enhanced granularity ontology rollup compared to the original ontology rollup pipeline.

Healthcare Scenario

The granularity enhancement takes place when collecting data from related tables of a structured EHR dataset and is a preprocessing of training/testing data before training a machine learning model. Our methods can be applied to any machine learning based applications when the structured EHR is encoded using prevalent medical vocabularies such as ICD-9, ICD-10, SNOMED, etc. Therefore, this method is both model agnostic (compatible with any type of machine learning such as neural networks, random forests, boosting methods, etc) and task agnostic (both classification or regression). For example, we have demonstrated the use of the granularity enhancement method on the task of mortality prediction and have observed performance improvement on the black group compared to the ontology rollup pipeline. Our method can be easily extended to other clinical tasks without much effort.

Operational Requirement

As our method focuses on the preprocessing of structured EHR data, we don't have specific requirements for computing resources such as OS, GPU, and RAM. We will provide light weighted Python classes and functions to enable the granularity enhancement process. All our Python codes will be distributed under the BSD 3 license. As a reference, the demonstration on the MIMIC-III dataset is generated under the below configurations:
Python: 3.8.16
Numpy: 1.23.1
Pandas: 1.5.3
Sklearn: 1.2.1
Scipy: 1.10.1
Shap: 0.41.0
LightGBM: 3.3.5

Generalizability Plan, Sustainability Plan & Implementation Requirements

As our method is both model-agnostic and task-agnostic, it can be widely applied to various application scenarios without much modification and cost. Up-to-date medical vocabularies are leveraged in our method to ensure the pipeline is digesting the most recent medical knowledge for granularity enhancement. We may need to invest time and human effort in tracking the medical vocabularies update (for example, the ICD-10 is supposed to update twice a year) and integrate the update into our Python codes. With existing model interpretability tools, the contribution of medical concepts, as well as the additional child concepts, can be easily obtained and visualized for clinicians or multidisciplinary experts to examine or review.

Lessons Learned

One challenge we faced is to decide if the underperformance of a model on a subgroup is caused by overfitting or underfitting. If the cause is overfitting, we may need to further rollup the feature set in order to control the dimensionality and if the cause of underfitting, we may need to add subconcepts to exhibit more details to the model. Our results show that the model can benefit from increasing granularity on essential features and therefore, one potential strategy to mitigate performance bias is to feed more details about the underperformed subgroup to increase the model's discriminative power on the subgroup.

Reference

[1] Fries JA, Steinberg E, Khattar S, Fleming SL, Posada J, Callahan A, Shah NH. Ontology-driven weak supervision for clinical entity classification in electronic health records. Nature communications. 2021 Apr 1;12(1):2017.
[2] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics. 2017 Oct 27;22(5):1589-604.
[3] Si Y, Du J, Li Z, Jiang X, Miller T, Wang F, Zheng WJ, Roberts K. Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review. Journal of biomedical informatics. 2021 Mar 1;115:103671.
[4] Solares JR, Raimondi FE, Zhu Y, Rahimian F, Canoy D, Tran J, Gomes AC, Payberah AH, Zottoli M, Nazarzadeh M, Conrad N. Deep learning for electronic health records: A comparative review of multiple deep neural architectures. Journal of biomedical informatics. 2020 Jan 1;101:103337.
[5] Bergquist T, Schaffter T, Yan Y, Yu T, Prosser J, Gao J, Chen G, Charzewski Ł, Nawalany Z, Brugere I, Retkute R. Evaluation of crowdsourced mortality prediction models as a framework for assessing AI in medicine. medRxiv. 2021 Jan 20:2021-01.
[6] Unlocking insights from real-world data to help pediatric COVID-19 patients. U.S. Department of Health & Human Services. Available from: https://www.medicalcountermeasures.gov/stories/pediatricchallengedata/
[7] Papatheodorou K, Banach M, Bekiari E, Rizzo M, Edmonds M. Complications of diabetes 2017. Journal of diabetes research. 2018 Mar 11;2018.