Fall risk prediction and validation in older adults

Leveraging electronic health records with machine learning

Dormosh, N.

Publication date
2023

Chapter

6

Impact of altering data granularity levels on predictive modelling: a case study of fall risk prediction in older persons

Noman Dormosh

Ameen Abu-Hanna

Nathalie van der Velde

Martijn C Schut

Studies in Health Technology and Informatics, Volume 270, p257-261, June 2020
DOI: 10.3233/shti200162

Abstract

Classification systems are widely used in medicine for knowledge representation. The hierarchical relationships between concepts in a classification system can be exploited in prediction models by looking for the optimal predictive granularity level. In this study, we used the Anatomical Therapeutic Chemical (ATC) classification system to cluster medications in the context of predicting medication-related falls in older persons. We compared the performance of fall risk prediction by describing medications at varying granularity levels of the ATC classification system. We found that the level of abstraction significantly affects the predictive performance in terms of both discrimination (measured by the receiver operating characteristic curve AUC-ROC) and calibration. An implication of these findings to the researchers is that data representation at different granularity levels can influence the predictive performance. The optimal granularity level can be determined by experimentation.

Introduction

Classification systems describe concepts and relationships between them. They are widely used for knowledge representation. The hierarchical relationship between concepts describes the relationship between a class and its sub-class. For example, the Anatomical Therapeutic Chemical (ATC) classification system (1) describes classes in terms of anatomy, therapy, and chemical compounds of medications and generalization specialization relationships between medications. Hence a medication can be described at various granularity levels. The choice of the appropriate level of abstraction to describe a class (here medication) depends on the task at hand.

When the task is prediction of an event (such as fall of a person), the problem can be stated as: at which level of granularity should medications be described in order to optimize the prediction of that event. This is not a trivial problem, as the error of a prediction model is governed by the bias-variance trade off (2): using a too coarse level of medications to predict falls may lead to parsimonious models with high bias, while using a too detailed level may lead to high variance. This necessitates search within the hierarchy to find the granularity level that leads to the best prediction model. In essence this is a feature selection strategy in prediction models.

Research studies have linked certain medications with an increased risk of falls in what is called fall-risk-increasing drugs (FRIDs) (3). FRIDs are grouped in fall risk prediction models to ease interpretation (4, 5). To exemplify the different granularity levels of these medications, consider “paroxetine”, this medication lies on the leaf level of the ATC hierarchy, and is a sub-class of a group called “selective serotonin reuptake inhibitors”. This group has a super-class named “antidepressants” which belongs to the “psychoanaleptics” class and further to the “nervous system” as the main anatomical level. In studies using medication as predictors, the decision on which level of abstraction to use them has often been arbitrary. In addition, little is known about the impact of medication abstraction on fall risk prediction performance.

The aim of this study was to understand the effect of describing medications at different granularity levels on the performance of fall risk prediction. We achieve this by comparing the predictive performance of models using various granularity levels of the ATC classification system.

Materials and methods

Study approach

In this cross-sectional study, we evaluated six prediction models which included the features pertaining to demographics, comorbidity; and home medications. Medications in the dataset are described at ATC level five (leaf level), which can be grouped at differ-

ent levels of the ATC classification system. We created six prediction models: ATC-5 a model that uses individual medications; ATC-1 to ATC-4 are models that group medications at the respective ATC level; and one model with only demographic and comorbidity features (No-Med).

Population, patient inclusion and clinical outcome

Data were obtained from the Medical Information Mart for Intensive Care (MIMIC-III, Version 1.4) (6) database. The MIMIC-III is a publicly available database containing deidentified data from more than 40,000 patients admitted to intensive care settings at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, MA, USA, between 2001 and 2012. The dataset comprises demographic elements such as age, gender and ethnicity, clinical elements including vital signs, laboratory data, diagnostic codes (in International Classification of Diseases 9th Edition, Clinical Modification [ICD-9-CM] format) and outcome values. Beside these it contains various types of clinical notes, including discharge summaries and nursing notes.

We included first admissions of older persons (≥65 years of age) with complete discharge summaries.

The clinical outcome was falls before hospital admission. Fallers were identified by the presence of an ICD-9 code that starts with "E88" which represent accidental falls, or the existence of any of the words (fall, fell, fallen) in the text of the reason of hospital admission field.

Data extraction and processing

The extracted demographic features were age, gender, ethnicity and marital status. We composed an algorithm to extract home medications and comorbidity features from the discharge summaries. The algorithm utilized RxNorm (7) and the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) in order to match the medications and the comorbidities in the text. Data processing involved the mapping of the extracted medications into ATC codes (level five) using the Observational Health Data Sciences and Informatics (OHDSI) (8) and obtaining all the super (parent) groups for each medication code.

Statistical analysis

We used the Least Absolute Shrinkage and Selection Operator (LASSO) with logistic regression to allow for feature selection as part of fitting the model and to avoid overfitting (9). LASSO trades off an increase in bias with a decrease in variance, by adding a penalty against model complexity. This penalty term (lambda) is usually chosen by cross-validation. The analysis was performed using the R programming language, ver-

sion 3.4 and data extraction was performed using Python. We used the implementation in the R glmnet package to perform LASSO regression.

We assessed the performance of the models in terms of discrimination (with the area under the receiver operating characteristic curve: AUC-ROC); calibration via calibration plots; and the relative goodness of fit with the Akaike Information Criterion (AIC). To internally validate the results, we applied bootstrapping (10) with 200 recommended repetitions and we calculated the optimism-corrected AUC-ROC. To get an honest assessment and to address LASSO instability of model performance all modelling steps (lambda-tuning) were repeated for each bootstrap sample. To test for statistical significance of the differences in the median between each of the models and the model with medications without grouping (ATC-5), we used Wilcoxon signed-rank test. A pvalue < 0.05 was considered to indicate statistical significance.

Results

The total number of patients was 46,520 in the MIMIC-III database. The number of patients which met the inclusion criteria was 23,034 (49.5%). The number of fallers was 1,721 (7.5% of the included patients). The number of different medications extracted from the discharge summaries was 939.

Figure 1 depicts the relationship between the AUC-ROC and the level of ATC grouping. It can be seen that the AUC-ROC initially increases when climbing up the hierarchy of the ATC, peaks at level two and then drops at the first level of the ATC.

The calibration performance of the models is shown in Figure 2. All the models showed acceptable calibration except the No-Med and ATC-4 models which had miscalibration. As shown in Table 1, the lowest AIC estimate (lower is better) was for the model ATC-2 followed by the model ATC-3. The number of selected features for these two models were 65 and 110 respectively.

Table 1: Akaike Information Criterion and the Number of Selected Features After Applying LASSO Regression

Models	AIC	Total features	Selected features
No-Med	-16,406.67	38	28
ATC-1	-16,615.84	52	37
ATC-2	-16,838.66	118	65
ATC-3	-16,655.52	212	110
ATC-4	-16,444.1	465	121
ATC-5	-16,369.3	977	196

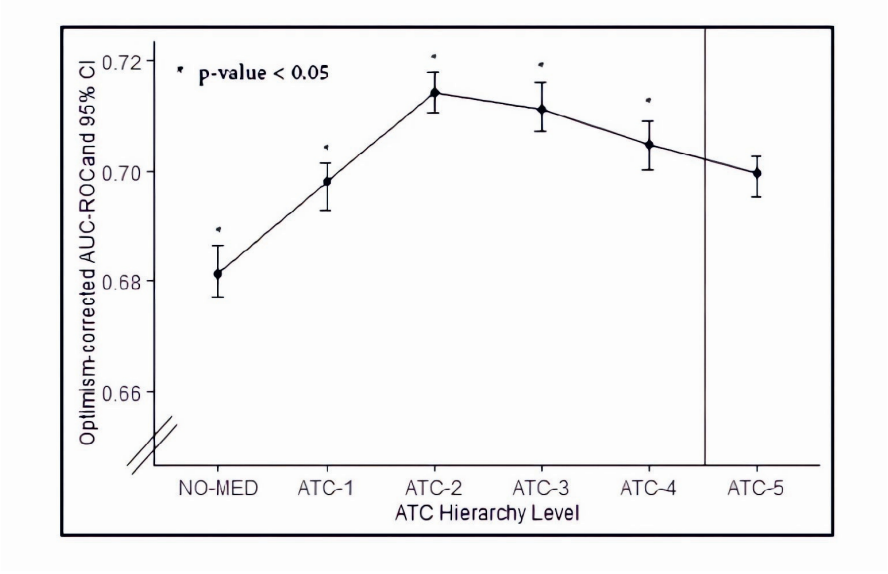

Figure 1: The relationship of AUC-ROC with different levels of grouping using the ATC. Note that the y-axis does not start from zero.

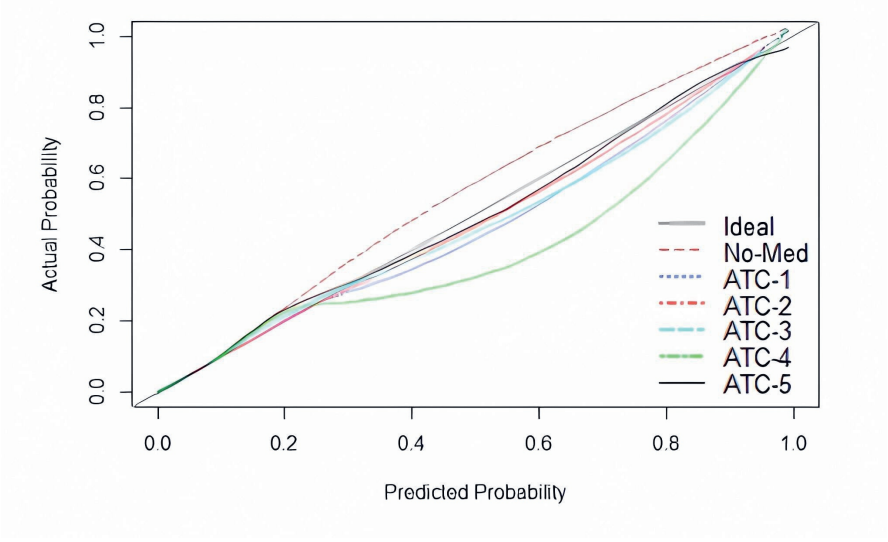

Figure 2: Bootstrap bias-corrected calibration curves of the models.

Discussion and conclusion

This study investigated the effect of leveraging hierarchical relationships of concepts in classification systems for predictive modelling. Our experiments indicate that significant improvement in predictive performance can be achieved by altering granularity levels. This is demonstrated by the ATC-2 model having the highest discrimination performance, acceptable calibration and the lowest AIC.

The strengths of our approach include a systematic evaluation of the performance of the prediction models and the use of bootstrapping to validate the results. The main limitation lies in the fact that the ICD 9 codes were generated for billing purposes at the end of the hospital stay and may not always represent falls before admission. Although the inspection of a random sample of the discharge summaries revealed that most of the patients fell before admission, there still is uncertainty in the definition of fall and the exact moment of its occurrence. This limitation may have affected the overall estimation of falls risk and prediction performance because we considered only home medications registered on admission and not during hospital stay. Furthermore, we could not validate the accuracy of the algorithms used to extract medications on admissions and comorbidities because of the lack of an annotated dataset. We however, believe these limitations should not qualitatively affect our findings as we evaluated the models under the same conditions.

The utilization of granularity in classification systems was also investigated in other domains. A recent study evaluated the impact of altering bacterial taxonomy on disease prediction (11). The authors found that describing bacteria at different granularity levels influenced the predictive performance. This finding corroborates what we observed in this work. Another recent study evaluated how different levels of data granularity of medications and diagnoses would affect prediction of diabetic kidney disease (DKD) prediction (12). The authors showed that different data granularity may not necessarily influence the predictive performance. However, this differs from the findings presented here and this discrepancy could be attributed to the differences in the domains on which the approach was evaluated.

On the basis of this work, we conclude that describing features at different levels of granularity can significantly affect the predictive performance. So far, we considered the same level of granularity to describe medications. A possible interesting extension to our work is to apply tree-lasso (13) or a (greedy-based) search strategy to select a subset of nodes in the tree-structured medication hierarchy. Another important work that warrants further investigation is validating our results on another dataset. Our findings have implications for researchers: we showed that prediction performance can be enhanced by altering the granularity of the data, and the optimal level of granularity can be obtained by experimentation.

References

[1] WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for atc classification and ddd assignment 2013. *Oslo*, 2012.

[2] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1 of *series in statistics*. Springer, 2001.

[3] M. de Vries, L. J. Seppala, J. G. Daams, E. M. M. van de Glind, T. Masud, N. van der Velde, Eugms Task, and Drugs Finish Group on Fall-Risk-Increasing. Fall-risk-increasing drugs: A systematic review and meta-analysis: I. cardiovascular drugs. *J Am Med Dir Assoc*, 19(4):371 e1–371 e9, 2018.

[4] P. Palumbo, L. Palmerini, S. Bandinelli, and L. Chiari. Fall risk assessment tools for elderly living in the community: Can we do better? *PLoS One*, 10(12):e0146247, 2015.

[5] M. L. Homer, N. P. Palmer, K. P. Fox, J. Armstrong, and K. D. Mandl. Predicting falls in people aged 65 years and older from insurance claims. *Am J Med*, 130(6):744 e17–744 e23, 2017.

[6] A. E. Johnson, T. J. Pollard, L. Shen, L. W. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark. Mimic-iii, a freely accessible critical care database. *Sci Data*, 3:160035, 2016.

[7] S. J. Nelson, K. Zeng, J. Kilbourne, T. Powell, and R. Moore. Normalized names for clinical drugs: Rxnorm at 6 years. *Journal of the American Medical Informatics Association*, 18(4):441–448, 2011.

[8] G. Hripcsak, J. D. Duke, N. H. Shah, C. G. Reich, V. Huser, M. J. Schuemie, M. A. Suchard, R. W. Park, I. C. Wong, P. R. Rijnbeek, J. van der Lei, N. Pratt, G. N. Noren, Y. C. Li, P. E. Stang, D. Madigan, and P. B. Ryan. Observational health data sciences and informatics (ohdsi): Opportunities for observational researchers. *Stud Health Technol Inform*, 216:574–8, 2015.

[9] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B-Methodological*, 58(1):267–288, 1996.

[10] Jr. Harrell, F. E., K. L. Lee, and D. B. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, 15(4):361–87, 1996.

[11] S. Bang, D. Yoo, S. J. Kim, S. Jhang, S. Cho, and H. Kim. Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. *Sci Rep*, 9(1):10189, 2019.

[12] X. Song, L. R. Waitman, Y. Hu, A. S. L. Yu, D. Robbins, and M. Liu. An exploration of ontology-based emr data abstraction for diabetic kidney disease prediction. *AMIA Jt Summits Transl Sci Proc*, 2019:704–713, 2019.

[13] J. Liu and J. Ye. Moreau-yosida regularization for grouped tree structure learning. *Advances in neural information processing systems*, 23, 2010.