Weak supervision for electronic phenotyping using electronic health records

Coauthor
Affiliation
Address
email

Coauthor
Affiliation
Address
email

Abstract

Electronic phenotyping is the process of identifying patients with the same latent health condition within a large population of patients based on their electronic health records (EHR). In order to generate discriminative machine learning models for phenotyping, there is a need to have large amounts of expert labeled patient sets with the target condition. Generating these large sets of patients is extremely time consuming as it requires groups of clinicians to manually review patient charts. Using weak supervision provided by imperfectly assigned labels and large amounts of patient data, it is possible to build models that perform as well as models built using manually curated training sets. In the open-source APHRODITE R package, we have operationalized two of the most popular approaches that leverage weak supervision and made them available to the Observational Health Data Sciences and Informatics (OHDSI) community and their common data model.

1 Introduction and background

Electronic phenotyping is the process of identifying patients with a medical condition or characteristic of interest based on records contained in an EHR system or clinical data repository using a defined set of data elements and logical expressions [1]. The goal of electronic phenotyping is to build patient cohorts or to ascertain outcomes in cohorts. The process of EHR-based phenotyping has evolved from using large sets of codes and rules manually curated by medical professionals to using machine-learning driven methodologies that can process large amounts of patient data with a wide variety of feature types and sources [2]. In order to avoid the work of manually creating sets of patients with the target phenotype to use with supervised learning approaches, we are turning to using weak supervision provided via cheaply acquired, imperfect labels.

The main idea of learning with imperfect labels leverages the result that imperfectly labeled data—used in larger amounts—can enable the learning of classifiers as good as those that can be learned from perfectly labeled data. An imperfect labeling procedure is one that assigns a wrong class label with a certain probability. Assuming a random classification noise (RCN) model [3], the probability of flipping labels is characterized by a parameter, called the classification error rate (τ). As derived in [4], [5] and used by [6], the amount of data needed for training a good model with noisy labels scales as $1/(1-2)^2$, where τ is the classification error rate. For τ=0 we have data with clean labels and τ=0.5 represents when the random flipping of labels destroys all signal in the data, making any learning impossible.

In this work, we discuss APHRODITE [7], an electronic phenotyping R-package which combines the ability of learning from imperfectly labeled data developed by Agaral et. al. as the XPRESS framework [6] and the Anchor learning framework for improving selected features in the phenotype models [8], for use with the OHDSI/OMOP CDM. The research contributions of this R package are that it allows researchers to build phenotyping models without needing hand labeled training data, allowing them to avoid relying too much on expertly curated gold standards. On the operational front, the contributions of this package are that it allows for the potential redistribution of locally validated phenotype models as well as the sharing of the workflows for learning phenotype models at sites of the OHDSI data network.

2 Framework evaluation

In order to evaluate the APHRODITE implementation of the XPRESS framework [6] and the Anchor learning framework [8], we have selected myocardial infarction as our phenotype of interest. Using the same manually curated evaluation gold-standard set from [6], we have built 'noisy' models using APHRODITE with 750, 1,500, and 10,000 cases and controls noisily identified. We then compare the performance of these models in identifying the gold standard patient set. Our first baseline comparison is how well the rule-based definition works (OMOP definition) and how the customly deployed XPRESS framework performed in [6]. We also evaluate the performance of APHRODITE models built following the Anchors and learn paradigm.

Table 1. Performance assessment of classifiers trained with noisy labeled training data

	Cases	Cont.	Acc.	Recall	PPV
OMOP definition [6]	94	94	0.87	*0.91*	0.84
XPRESS [6]	94	94	0.89	0.93	0.86
APHRODITE (750)	94	94	0.91	0.93	0.90
APHRODITE (1,500)	94	94	0.92	0.93	**0.91**
APHRODITE (10,000)	94	94	0.92	0.94	**0.91**
APHRODITE (Anchors)	94	94	0.92	**0.97**	0.89
APHRODITE (Anchors + features mod)	94	94	**0.93**	0.96	**0.91**

As Table 1 clearly shows, the machine learning models, rows 2 to 7, perform equally or better than the rule-based definitions (row 1). The most important result here is that the models built by APHRODITE perform equally or better than XPRESS, showcasing the power of our framework as it does not need to be implemented from scratch at new sites, or have the data fit the original XPRESS specifications. A mode detailed and through evaluation can be found in [7].

3 Maximizing Impact

With the formation of partnerships like PCORnet (the National Patient-Centered Clinical Research Network) which uses the Mini-Sentinel Common Data Model (CDM), and the OHDSI data network which uses the OMOP CDM, there is an increasing need for accurate and fast methods for electronic phenotyping. The patient network available in the common OHDSI CDM includes over 100 databases, both clinical and claims, totaling over 650 million patients. APHRODITE was developed to tap into this data network and alleviate the need to implement XPRESS or the anchors and learn framework for site-specific studies. APHRODITE has been used in the following scenarios: successfully phenotype patients with familial hypercholesterolemia in data from two separate institutions [9], identify cases of metastatic prostate cancer [10], and to study patient comorbidities [11], just to name a few. Addressing one of the biggest issues of machine learning in healthcare, model portability across multiple institutions, a recent study has shown that APHRODITE models show promise with regards to portability across institutions [12]. This study develops and shares models across two USA institutions and one international with very promising results, showcasing the power of noisy labeling for electronic phenotyping.

References

[1] "NIH Collaboratory Rethinking Clinical Trials - The Living Textbook - Rethinking Clinical Trials," *Rethinking Clinical Trials*. [Online]. Available: https://rethinkingclinicaltrials.org/. [Accessed: 06-Sep-2019]

[2] J. M. Banda, M. Seneviratne, T. Hernandez-Boussard, and N. H. Shah, "Advances in Electronic Phenotyping: From Rule-Based Definitions to Machine Learning Models," *Annu Rev Biomed Data Sci*, vol. 1, pp. 53–68, Jul. 2018.

[3] P. M. Long and R. A. Servedio, "Random classification noise defeats all convex potential boosters," *Mach. Learn.*, vol. 78, no. 3, pp. 287–304, Mar. 2010.

[4] J. A. Aslam and S. E. Decatur, "On the sample complexity of noise-tolerant learning," *Inf. Process. Lett.*, vol. 57, no. 4, pp. 189–195, Feb. 1996.

[5] H. U. Simon, "General Bounds on the Number of Examples Needed for Learning Probabilistic Concepts," *J. Comput. System Sci.*, vol. 52, no. 2, pp. 239–254, Apr. 1996.

[6] V. Agarwal *et al.*, "Learning statistical models of phenotypes using noisy labeled training data," *J. Am. Med. Inform. Assoc.*, vol. 23, no. 6, pp. 1166–1173, Nov. 2016.

[7] J. M. Banda, Y. Halpern, D. Sontag, and N. H. Shah, "Electronic phenotyping with APHRODITE and the Observational Health Sciences and Informatics (OHDSI) data network," *AMIA Jt Summits Transl Sci Proc*, vol. 2017, pp. 48–57, Jul. 2017.

[8] Y. Halpern, S. Horng, Y. Choi, and D. Sontag, "Electronic medical record phenotyping using the anchor and learn framework," *J. Am. Med. Inform. Assoc.*, vol. 23, no. 4, pp. 731–740, Jul. 2016.

[9] J. M. Banda *et al.*, "Finding missed cases of familial hypercholesterolemia in health systems using machine learning," *NPJ Digit Med*, vol. 2, p. 23, Apr. 2019.

[10] M. G. Seneviratne, J. M. Banda, J. D. Brooks, N. H. Shah, and T. M. Hernandez-Boussard, "Identifying Cases of Metastatic Prostate Cancer Using Machine Learning on Electronic Health Records," *AMIA Annu. Symp. Proc.*, vol. 2018, pp. 1498–1504, Dec. 2018.

[11] A. Y. Ling, E. Alsentzer, J. Chen, J. M. Banda, S. Tamang, and E. Minty, "Scalable Electronic Phenotyping For Studying Patient Comorbidities," *AMIA Annu. Symp. Proc.*, vol. 2018, pp. 740–749, Dec. 2018.

[12] M. Kashyap *et al.*, "Development and Validation of Phenotype Classifiers across Multiple Sites in the Observational Health Sciences and Informatics (OHDSI) Network." [Online]. Available: http://dx.doi.org/10.1101/673418