

An Algorithm for Mapping Local Measurement Concepts into OMOP (Observational Medical Outcomes Partnership) Vocabulary Using Logical Observation Identifiers Names and Codes (LOINC®) Terminology

Ergin Soysal, MD, PhD, Jingqi Wang, MS, Min Jiang, MS, Hua Xu, PhD
School of Biomedical Informatics, University of Texas Health Science Center at Houston,
Houston, Texas

Abstract

We present a software utilizing machine learning techniques for translating locally used measurement concept names and codes in an electronic health record data into corresponding Logical Observation Identifiers Names and Codes (LOINC) terms. LOINC is a part of Observational Medical Outcomes Partnership (OMOP) Vocabulary in measurement domain. Algorithm developed using a manually encoded large laboratory names data of 21 thousand laboratory names, units, value types and normal ranges as well as specimen sources whenever available.

Introduction

It's a common problem to map locally used laboratory test names and codes from electronic health record (EHR) data into the globally accepted standard LOINC (Logical Observation Identifiers Names and Codes) counterparts. Because of diversity of laboratory tests, this even becomes a difficult task for human users. Overlapping semantics of LOINC concepts makes the task even more difficult.

Methods

Laboratory test names over 21 thousand used in outpatient clinics of Memorial Hermann Hospitals were collected and organized by specimen sources, value types and ranges, and units. Tests were manually assigned corresponding LOINC codes by 2 domain experts with an agreement of 96% using a web based software developed for this purpose.

Using LOINC concepts, a master measurement index was prepared. This index was enriched by synonyms from UMLS using the terms with the same CUIs from other terminologies like SNOMED CT and MeSH. An abbreviation vocabulary was also used which was created during manual review of laboratory names. The annotated dataset divided into two equal sets for development and test of the model. At first, for each laboratory test name from the development set, a group of 20 candidate LOINC concepts were selected from the measurement index based on the TF-IDF score rankings. These 20 candidates were covering the correct LOINC codes for the laboratory test name in 90% of the cases. Then an SVM model was trained using these candidates with respect to gold standard annotation as a binary label and laboratory test name, candidate concept name, TF-IDF rank and other available test properties as model features.

Results

The model was tested using the spare test dataset over 10 thousand laboratory names. The system achieved a performance of 81% accuracy in encoding laboratory test names into corresponding LOINC Codes, which then converted into corresponding OMOP vocabulary concept ids.

Conclusion

We present a software using machine learning algorithms, which can be helpful in mapping EHR application specific non-standard concepts into OMOP vocabulary concepts with an acceptable success rate. Because of the nature and size of laboratory name entries, manual mapping can be overwhelming and time consuming. The software can be a valuable tool in this conversion process.