Efficient automated mapping of internal source codes to OMOP CDM concepts

IOMED Medical Solutions

Automated Concept Mapping System for OMOP using Vector Representations and Cross-hospital Mapping

The searching based manual assignment of standard concepts in Observational Medical Outcomes Partnership (OMOP) to local source codes is a time-consuming and error-prone process. Although the OHDSI-developed tool USAGI was designed for this procedure, the non-English language data and OMOP integration limitations hinder efficient mapping

Methods

We propose a method that leverages existing mappings from other hospitals, enabling the efficient scaling of a single mapping across all others. The mapping process is automated by computing vector representations of source code texts, which capture the relevant syntactic and semantic features, ensuring that similar records are grouped closely in the vector space. Consequently, assigning concept IDs becomes a matter of performing **similarity searches** within the vector space.





Electronic Health Records (EHRs) of multiple hospitals Source descriptions are **preprocessed** and later vectorized using an Natural Language Processing (NLP) model

With vector similarity calculation, the most similar records are identified to compute an **automatic mapping**

Clinical specialists validate model results. We show them the top 5 concepts

Update the *_concept_id OMOP CDM tables fields



- Data sourced from **11 distinct hospitals**
- A total of **79.928 unique source codes**

72%

Number of correct mappings executed by the model

88%

In our use case, shows how likely it is for the correct mapping to be amongst the ones provided to the clinical specialist





Martina Carrés martina.carres@iomed.es

78.07% reduction in mapping

validation time achieved

Gabriel Maeztu Mónica Arrúe gabriel.maeztu@iomed.es monica.arrue@iomed.es

