Classifier Chains for LOINC Transcoding

Théodore Michel-Picque, Sandra Bringay, Pascal Poncelet, Namrata Patel, Guilhem Mayoral

Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240654

Classifier Chains for LOINC Transcoding

Théodore MICHEL-PICQUE [a,c,1], Sandra BRINGAY [a,b], Pascal PONCELET [a]
Namrata PATEL [b,c] and Guilhem MAYORAL [c]
[a] *LIRMM UMR 5506, University of Montpellier, CNRS, Montpellier, France*
[b] *AMIS, Paul-Valery University, Montpellier, France*
[c] *Onaos, Montpellier, France*
ORCiD ID: Théodore MICHEL–PICQUE https://orcid.org/0009-0006-9538-2153, Sandra BRINGAY https://orcid.org/0000-0002-2830-3666, Pascal PONCELET https://orcid.org/0000-0002-8277-3490, Namrata PATEL https://orcid.org/0000-0002-7923-4840, Guilhem MAYORAL https://orcid.org/0000-0001-9801-2090

Abstract. Purpose: Mapping clinical observations and medical test results into the standardized vocabulary LOINC is a prerequisite for exchanging clinical data between health information systems and ensuring efficient interoperability. Methods: We present a comparison of three approaches for LOINC transcoding applied to French data collected from real-world settings. These approaches include both a state-of-the-art language model approach and a classifier chains approach. Results: Our study demonstrates that we successfully improve the performance of the baselines using the classifier chains approach and compete effectively with state-of-the-art language models. Conclusions: Our approach proves to be efficient, cost-effective despite reproducibility challenges and potential for future optimizations and dataset testing.

Keywords. LOINC Transcoding, Interoperability, Clinical Data Standardization, Language model

1. Introduction

LOINC[2] [1] (Logical Observation Identifiers Names and Codes) is an internationally recognized reference terminology employed since the late 1990s for identifying and labeling clinical observations and medical test results. It establishes a standardized coding system for laboratory tests, clinical measurements, and various health observations. Each LOINC code comprises a combination of numbers and alphanumeric characters, precisely delineating the nature of the observation or test. LOINC enjoys widespread adoption globally within healthcare facilities, facilitating the seamless exchange of clinical data between health information systems and ensuring efficient interoperability. This article outlines a LOINC transcoding procedure in French[3], aiming to establish an automated mapping between locally cataloged laboratory examinations

[1] Corresponding Author: Théodore MICHEL-PICQUE, LIRMM, 161 Rue Ada, 34095 Montpellier France; E-mail: tmichelpicque@lirmm.fr.

[2] https://loinc.org.

[3] In France, the adoption of this standard has been mandated within French laboratories since the Ségur du Numérique in 2016, as outlined in a roadmap devised by the French Ministry of Solidarity and Health.

and their corresponding LOINC codes. In the context of LOINC code transcoding, previous research has demonstrated the utility and relevance of manual or semi-manual transcoding tools, of- ten managed by medical experts [2]. However, due to the time-intensive nature of the task and the impracticality of delegating it to individuals with less training than medical experts, this method is not currently the preferred approach. With the widespread adoption of machine learning, new advancements in this field have already been made. For instance, Kelly et al. [3] developed innovative encoding methodologies to vectorize free-text laboratory data and evaluated the performance of logistic regression, random forest, and K-nearest neighbors machine learning classifiers for LOINC transcoding. Tu et al. [4] leverage contextual embedding from pre-trained T5 models and propose a two- stage fine-tuning strategy based on contrastive learning. While these methods are highly effective, those based on language models present challenges in interpretation and raise ethical and bias considerations due to their potential to inherit societal biases present in the data, leading to biased or unfair outcomes. Additionally, fine-tuning these models for specific tasks can be expensive and may require a large volume of data. Finally, these models often overlook the interconnections between different parts of the LOINC code. In this article, drawing inspiration from the work of Read et al. [5] on classifier chains, we do not approach transcoding as a simple task of multi-label classification. Instead, we consider the correlations among components in the label, proposing a novel classifier chains method that models these component correlations. The article makes a twofold contribution. Through extensive experiments aimed at standardizing French laboratory data according to the International LOINC format, we achieve two key outcomes: 1) we demonstrate the effectiveness of the state-of-the-art in LOINC code mapping based on language models, and 2) we propose a cost-effective alternative that, based on the classifier chains method [5], yields slightly better results by enabling the identification of the correct sequence of components in the labels. The remainder of the article is organized as follows: Section 2 outlines the methods and implementations, while Section 3 provides details on the conducted experiments and presents the results. Finally, we conclude by discussing our future work.

2. Methods

2.1. Task description

Transcoding involves predicting the LOINC code, a numerical code like '42254-3,' unique to each examination. As illustrated in Figure 1, this code has a textual equivalent known as 'Complete LOINC' (e.g., 'Noyau anticorps [Présence/Seuil] ; Serum ; Qualitatif,' translated into English as 'Antibody core [Presence/threshold]; Serum; Qualitative') that can be broken down into components, such as System (e.g., 'Serum'), Component (e.g., 'Antibody core'), Scale (e.g., 'Qualitatif'), etc. Considering a dataset D comprising examination results from French laboratories, each examination result xi is represented by a feature vector used for prediction. An illustrative example of such an entry is provided in Table 1. The prediction target for each examination, denoted as yi, is a vector defined as yi = Si, Pi, Sc, Laci, Lcompi, Cli, where Si represents the System, Pi the Property, Sci the Scale, Laci the Analyte core, Lcompi the Component, and Cli the Complete LOINC.

Table 1. Illustrative examples of an entry xi

Customer code	Local label	Synonyms	NABM[4]	Scale
AC	AC ANTINUCLEAR	Anti-nuclear autoantibodies	324	Numerical
Technique	**Chapter code**	**System**	**Sample tube**	**Units**
Smear	Calculation	Serum	M4RT	g/l

As illustrated in Figure 1, the hierarchical relationship between *Cl*, *Lcomp*, and *Lac* can be represented through an inclusion hierarchy, where $Lac \subset Lcomp \subset Cl$. This indicates that *Lac* is included in *Lcomp* and *Lcomp* is included in *Cl*. Additionally, it should be noted that *S*, *Sc*, and *P* are also included within *Cl*, further emphasizing the extensive and nested structure of these relationships within the hierarchy. The objective is to develop a prediction model $f: X \rightarrow Y$ to accurately map the input feature vectors X to the target vectors Y, considering the hierarchical structure of the LOINC components. The model aims to optimize a loss function $L\ (f(x_i), y_i)$ over the dataset D, where $f(x_i)$ represents the model's prediction for input x_i, and y_i is the actual target vector. The selection of the loss function should reflect the accuracy of the prediction within the hierarchical structure of the LOINC codes.

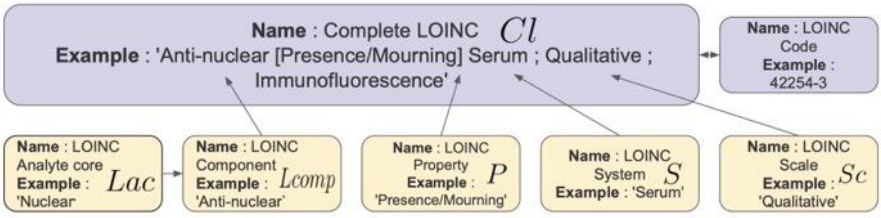

Figure 1. Hierarchical organization of LOINC attributes

2.2. Models

Below, we offer a comprehensive overview of the three models we tested.

- **Baseline**: In addressing a multi-class problem, we utilize the original structure of the LOINC dataset without considering the hierarchy between the components to predict the final target Cl_i.
- **Text-Based**: We address a sentence classification problem by predicting the final target Cl_i based on the concatenation of all input features.
- **Classifier chains**: Instead of predicting only the final target Cl_i, we begin with intermediate targets and incorporate them as features for predicting the next target, ultimately leading to the prediction of the final target Cl_i.

3. Experiments and Results

3.1. Experiments

Experiments were conducted on a real dataset collected from 99 French laboratories, comprising 162,678 entries for 11,216 unique LOINC codes. We refined this distribution by retaining a code only if it appears at least 10 times, resulting in a final dataset of 139,235 entries and 2,463 LOINC codes. We employed different implementations for each method.

Baseline: We encoded the features with a OneHotEncoder and the target with a LabelEncoder. With this preprocessing, we were able to train SGD and XGBoost models to predict the final target Cl_i by approaching the problem as a multi-class classification. The default hyperparameters were used for both models.
Text-Based: We began by normalizing and concatenating all the textual components of laboratory entries. For the target, we use the Complete LOINC *Cl*. Subsequently, we used CAMEMBERT [6], a French general-purpose language model, and DrBERT [7], a French biomedical language model. Both are state-of-the-art language models based on RoBERTa [8]. We trained them for 20 epochs, using a batch size of 64 and the AdamW optimizer.
Classifier chains: Initially, we encoded each target with a different instance of LabelEncoder and initialized a classifier for each step of our algorithm. We then encoded our features with a OneHotEncoder, repeating the operation each time we incorporated the previous target (predicted in our input with the same preprocessing). We use the same classifiers and the same hyperparameters as in the baseline to facilitate a comparison of our approaches.

3.2. Results

To assess the performance of our approaches, we employed Precision, Recall, and F1-score as metrics. All metrics are weighted to consider the heterogeneity of the data distribution. The results are presented in Table 2.

Table 2. Comparison of the results of our different approaches

Method	Precision (Weighted) (%)	Recall (Weighted) (%)	F1-Score (Weighted) (%)
Baseline SGD	0.85	0.83	0.84
Baseline XGboost	0.85	0.75	0.80
Text-based CamemBERT	0.84	0.80	0.82
Text-based DrBERT	0.81	0.81	0.81
Classifier chains SGD	**0.88** (+0.03)	**0.87** (+0.04)	**0.87** (0.03)
Classifier chains XGBoost	0.81 (-0.04)	0.82 (+0.07)	0.82 (+0.02)

The results of our comparison reveal several findings. All the methods show good results, with the lowest F1-Score recorded for the Baseline XGBoost at 0.80. The performances of the two text-based models are very similar, suggesting that CAMEMBERT's specialization for the French language and DrBERT's specialization for medical data are significant enough to bring about a meaningful difference. The best method is the classifier chains approach using SGD, achieving an F1-Score of 0.87, which outperforms the others in the two-remaining metrics (Precision of 0.88 and Recall of 0.87).

4. Discussions and Conclusions

In this work, we conducted extensive experiments for LOINC transcoding into French. We compared basic machine learning approaches, a state-of-the-art approach based on language models, and a novel method utilizing classifier chains, which, in an original manner, considers the dependency links among the label components. The classifier chains approach using SGD emerged as the most effective, achieving an F1-Score of 0.87. Initial findings from real-world data are promising, signaling the necessity for additional experiments and further exploration. However, comparing the performance with other studies may be challenging given the diversity in the number of covered LOINC codes. It is also conceivable that language models could yield improved performance with further optimizations and fine-tuning. Finally, our classifier chains model may face challenges with data shift on another dataset. As a next step, we plan to test this approach on various datasets, including the MIMIC dataset.

Acknowledgments

This material is based upon work supported by the ANRT (Association nationale de la recherche et de la technologie) with a CIFRE fellowship granted to MICHEL-PICQUE Théodore.

References

[1] Rocha RA, Huff SM. Coupling vocabularies and data structures: lessons from LOINC. In: Proceedings of the AMIA Annual Fall Symposium, 1996:90-94.
[2] Nikiema JN, Vianney RG. Aligning an Interface Terminology to the Logical Observation Identifiers Names and Codes (loinc®). Journal of the American Medical Informatics Association Open. 2021 Jun 12;4(2).
[3] Kelly J, Wang C, Zhang J, Das S, Ren A, Warnekar P. Automated mapping of real-world oncology laboratory data to LOINC. In: Proceedings of the AMIA Annual Fall Symposium, 2021:611–620.
[4] Tu T, Loreaux E, Chesley E, Lelkes AD, Gamble P, Bellaiche M, et al. Automated Loinc Standardization Using pre-trained Large Language Models. In: Proceedings of the 2nd Machine Learning for Health symposium. 2022. p. 343–55.
[5] Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. Machine learning. 2011; 85:333–359.
[6] Martin L, Muller B, Ortiz Suárez PJ, Dupont Y, Romary L, de la Clergerie É, et al. CamemBERT: A tasty french language model. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, 7203-7219.
[7] Labrak Y, Bazoge A, Dufour R, Rouvier M, Morin E, Daille B, et al. DrBERT: A robust pre-trained model in French for biomedical and clinical domains. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023, 16207-16221.
[8] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692v1, 2019.