Crowdsourcing for ICD10 Code to Concept Relationships

Patrice Seyed[1], Evan Patton[2], and Michael Subotin[1]

[1] 3M HIS, Silver Spring, MD
[2] Dept. of Computer Science, RPI, Troy, NY

Abstract. In this work we leverage crowdsourcing in connection with machine learning techniques to validate candidate ICD10 Code to UMLS concept relationships that we generate. Our immediate use is in natural language understanding and machine learning approaches to automatically code electronic health record documents with ICD codes. Beyond auto-coding, the relationships will aid a wide variety of future medical applications, such as terminology-driven search in support of smart medical assistants.

1 Introduction

One of the current challenges in the domain of clinical informatics is improving upon the quality and coverage of equivalence and compositional relationships between medical terminologies found in the Unified Medical Language System (UMLS) [2]. Crowdsourcing has proven a vital resource for this sort of annotation, for example, in an effort to identify relations in clinical documents [1]. In this work we aim to improve the International Classification of Disease, Clinical Modification (ICD10-CM) auto-coding of electronic health records (EHRs) by leveraging crowdsourcing in connection with machine learning techniques to validate generated relationships linking ICD10 CM codes to concepts of terminologies included in UMLS. While there exists effort elsewhere for generating such relationships [4], the results focus on equivalence relationships and driven by lexical matching, whereas our work covers EHR corpus-driven candidate generation, and includes taxonomic, equivalence, and other direct relationships (e.g., 'finding site', 'diagnostic procedure for') that directly hold between concepts and clinical situations represented as ICD codes.

2 Workflow

Initially, EHRs manually annotated with gold standard ICD-10 CM Codes are also annotated for UMLS concepts, and the code-concept pairs are processed for pointwise mutual information (PMI) across a corpus [3]. We consider PMI scores, as well as concept counts and pair count as thresholds for determining the set of pairs for submitting for crowdsourcing using Amazon's Mechanical Turk. The use of similarity (e.g., lexical, semantic) measurements is planned to

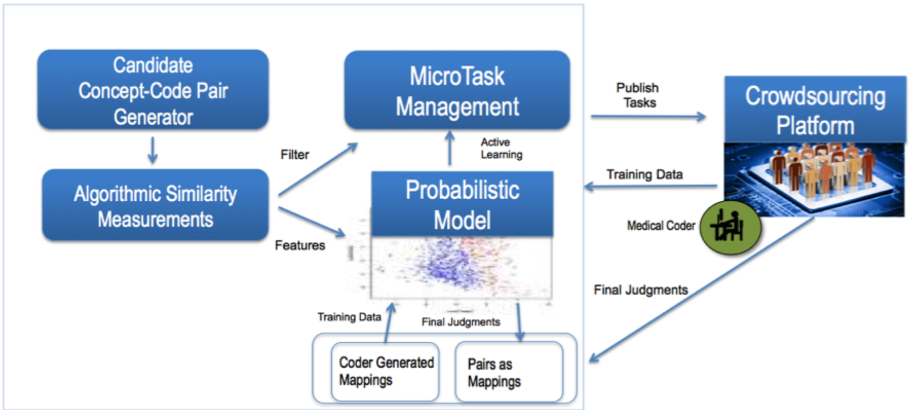

Fig. 1. Overall Workflow

assist filter candidates, and as features for training a machine learning model that will identify the relationship (if it exists) between an arbitrary ICD10 CM code and concept.

Once we have trained our probabilistic model with crowdsourcing results, we use the results to predict whether the candidate pairs that have not yet been crowd-sourced are valid or not (using PMI and other measures as features, and results of the crowd as labeled data). In the initial phase, we leverage subject matter experts' knowledge to validate the crowdsourced judgments. Thus the crowdsourcing data is used for two purposes: training the probabilistic model and for final judgments. As similarity measures are developed further within the project, they can be used as additional features to identify candidates to crowd source or serve as relationships. Table 1 shows three pairs examples. The overall workflow is illustrated in Figure 1.

Table 1. Example Candidate Code-Concept Pairs

CM Code	Concept	Concept Type	PMI	Pair Count	Concept Count
Tinnitus, unspecified ear (H93.19)	Rinne test (C0278245)	Diagnostic Procedure	2.26	11	163
Localized swelling, mass and lump, neck (R22.1)	Stereotactic Imaging (C0729296)	Diagnostic Procedure	1.80	21	433
Aortic aneurysm of unspecific site, without ruptured (I71.9)	Repair of Aneurysm (C0189661)	Preventative Procedure	2.13	11	559

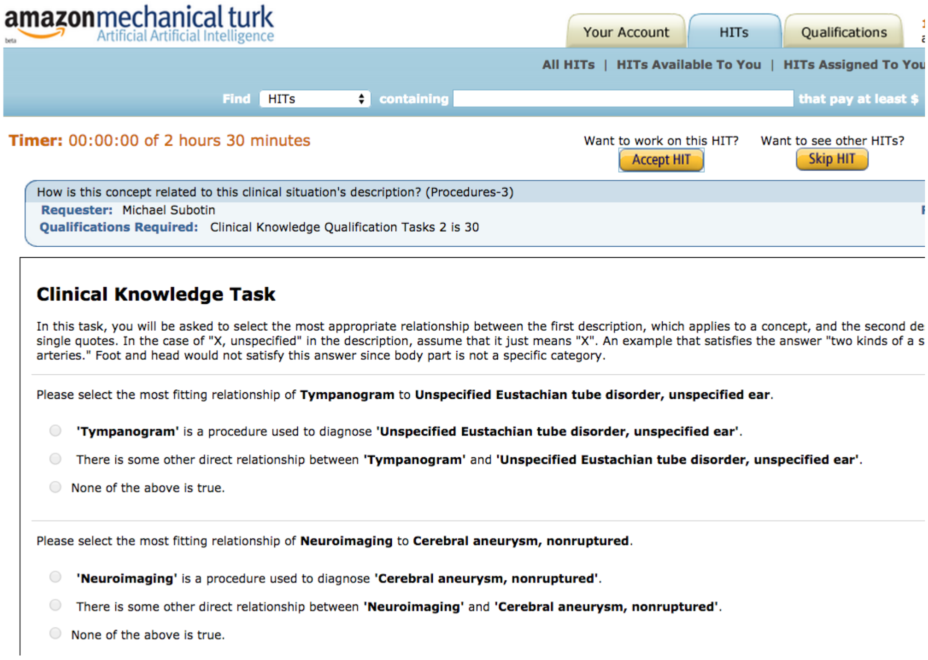

Fig. 2. Example Questions Published on Mechnical Turk.

3 Designing Questions for the Crowd

In order to determine how best to formulate natural language questions to ask as to the direct relationship between a code (i.e., clinical situation) and a concept, we consider the UMLS semantic type, based primarily on the following: Disorders, Body Parts, Procedures, and Findings. For example: if the concept is a disorder, the question options pertain to the taxonomic relation “isa”; if the concept is a body part, the question pertain to whether it is the finding site of the disorder; if the concept is a diagnostic procedure, the questions pertain to whether the procedure is used to diagnose the disorder. Note then, that the relationships are both taxonomic, compositional, and other direct relations, therefore supporting structured knowledge source applications. Example questions posed to Mechanical Turk workers are given in Figure 2.

4 Evaluation And Future Steps

For evaluating our results, we consider majority for confirming a specific answer, and consensus for confirming negation (i.e., none of the above answer) is accurate. We are in the process of adjusting parameters (pay, questions per task, qualification questions, assignment duration, auto-approval). Our next steps include performing analyses for evaluation techniques for workers, answers, and question quality. This is useful since disagreement is oftimes signal and not noise [1]. We aim to increase the volume of results by increasing pay and comparing results against another crowsourcing platform, CrowdFlower. As for utility, we will investigate which code-to-concept pairs are directly useful in our rule-based systems, and which are useful primarily for machine learning approaches for auto-coding.

5 Vision and Impact

The improved ICD-10 to UMLS relationships generated by our crowd-sourcing approach will result in less noisy and more robust structured knowledge. It will enable the use of these relationships are rules as well as evidence for auto-coding ICD 10 CM. Also, such knowledge will support future medical applications aimed at aiding practitioners and patients alike. Use of knowledge representation for building medical expert systems for diagnosis has been well explored (for a review of techniques, see[6]). The number of structured platforms for patient-oriented smart medical assistants is also growing, for example [5, 7]. All of these applications using structured knowledge approaches stand to benefit from our crowd-sourced relationships.

References

1. Aroyo, L., Welty, C.: Measuring crowd truth for medical relation extraction. In: 2013 AAAI Fall Symposium Series (2013)
2. Bodenreider, O.: The unified medical language system (umls): integrating biomedical terminology. Nucleic acids research 32(suppl 1), D267–D270 (2004)
3. Bouma, G.: Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL pp. 31–40 (2009)
4. Campbell, J.R., Brear, H., Scichilone, R., White, S., Giannangelo, K., Carlsen, B., Solbrig, H.R., Fung, K.W.: Semantic interoperation and electronic health records: context sensitive mapping from snomed ct to icd-10. In: MedInfo. pp. 603–607 (2013)
5. Lewandowski, J., Arochena, H.E., Naguib, R.N., Chao, K.M., Garcia-Perez, A.: Logic-centered architecture for ubiquitous health monitoring. Biomedical and Health Informatics, IEEE Journal of 18(5), 1525–1532 (2014)
6. Liao, S.H.: Expert system methodologies and applications—a decade review from 1995 to 2004. Expert systems with applications 28(1), 93–103 (2005)
7. Patton, E.W., McGuinness, D.L.: Toward next generation integrative semantic health information assistants. In: Proc. AAAI Winter Symp. on Expandings the Boundaries of Health Informatics using Artificial Intelligence (2014)