

Ontology-based Normalization for Disease-Lab test Relation Extraction

Yaoyun Zhang, Jingqi Wang, Cui Tao, Hua Xu

School of Biomedical Informatics
University of Texas at Houston
Houston, USA

{Yaoyun.Zhang, Jingqi.Wang, Cui.Tao, Hua.Xu}@uth.tmc.edu

Abstract—This poster describes our preliminary work on ontology-based normalization for diseases and lab tests, as a fundamental step toward disease-lab test relation extraction. Multiple ontologies are leveraged for this aim. Specifically, diseases and lab tests are first extracted and mapped to the Concept Unique Identifier (CUI) of the Unified Medical Language System (UMLS) by MetaMap. Codes of International Classification of Diseases, Version 9 – Clinical Modification (ICD-9CM) are then employed to further normalize diseases; while the Logical Observation Identifiers Names and Codes (LOINC) are used to normalize lab tests.

Keywords—ontology-based normalization; disease normalization; lab test normalization; relation extraction

I. INTRODUCTION

Disease-labtest relation extraction plays an important role in various medical applications such as clinical decision-support systems and phenotype information extraction. However, mentions of diseases and labtests in text contains diverse non-standard variations. Those variants need to be normalized into standard codes first to facilitate more universal computational applications. This poster describes our preliminary work on ontology-based normalization for diseases and lab tests, as a fundamental step toward disease-lab test relation extraction. Three existing standard ontologies UMLS[1], ICD-9CM[3] and LOINC[4] are leveraged for this aim.

II. ONTOLOGY OVERVIEW

Overview

A. UMLS

UMLS[1] is a thesaurus re-organizing many controlled vocabularies in the biomedical sciences. It provides a mapping structure among various vocabularies and serves as a comprehensive ontology of biomedical concepts. For each concept in UMLS, a synonym list consisting of terms from multiple vocabularies is collected. For example, “Diabetes Mellitus” and “dm” are synonyms of the same CUI C0011849. Various ontological relations between CUIs are defined, such as “isa”, “broader”, and “sibling”, etc.

B. ICD-9CM

The International Statistical Classification of Diseases and Related Health Problems (ICD) is the international "standard diagnostic tool for epidemiology, health management and clinical purposes"[3]. The ICD provides a hierarchical system of diagnostic codes for classifying diseases. Major categories are designed to include a set of similar diseases as sub-categories. For example, “(050.0) Variola major” is a subcategory of “(050) Smallpox”. Health conditions can be mapped corresponding generic categories or more specific sub-categories.

C. LOINC

The Logical Observation Identifier Names and Codes (LOINC) is the only publicly available universal standard for laboratory test codes and names[4 5]. The current version of the LOINC code set (released in June 2014) contains 73,889 terms for lab tests, measurements and clinical observations. Lab tests are organized hierarchically into 14 top classes including “Microbiology”, “Blood Bank”, etc.

III. NORMALIZATION METHOD

The original mention of disease/lab test is first recognized and mapped to UMLS concepts by MetaMap[2]. If ICD-9CM/LOINC is among the sources of terms for the mapped UMLS concept, then the mention can be normalized to the corresponding code and name from ICD-9CM/LOINC directly. If not, for disease, if SNOWMED CT is one source of terms for the UMLS concept, the corresponding SNOWMED CT concept can be mapped to ICD-9CM code by the rule-based mapping provided by NIH[6]. For lab test, the RELMA software is employed to map labtest to LOINC code and name[7]. Fig. 1 and Fig. 2 illustrate the workflow of disease normalization and lab test normalization, respectively.

Fig. 1. Diagram of Disease Normalization

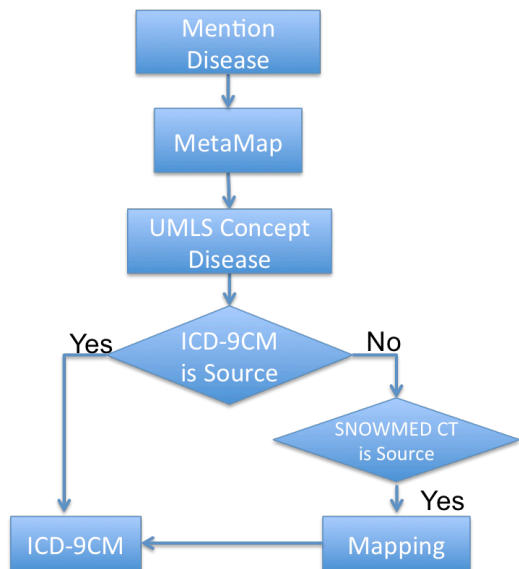
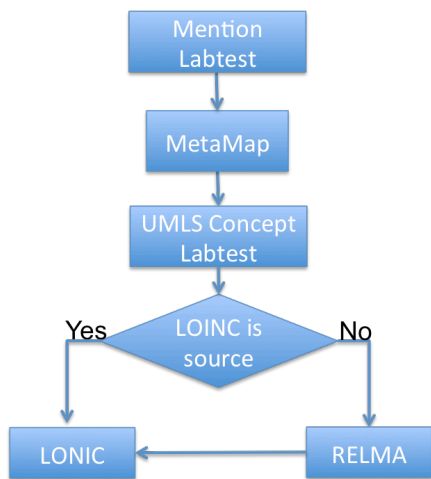


Fig. 2. Diagram of Lab test Normalization



IV. NORMALIZATION RESULTS

Multiple variations of diseases and lab tests are normalized into standard codes following the workflow in Fig 1 and Fig 2. Fig. 3 and Fig. 4 show the examples of disease normalization and lab test normalization, respectively. The original mentions are mapped to UMLS concept first, and then to ICD-9 CM and LOINC code.

Fig. 3. Example of Disease Normalization

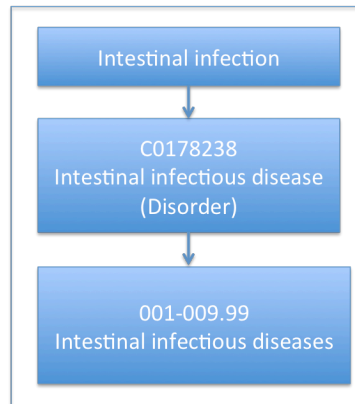
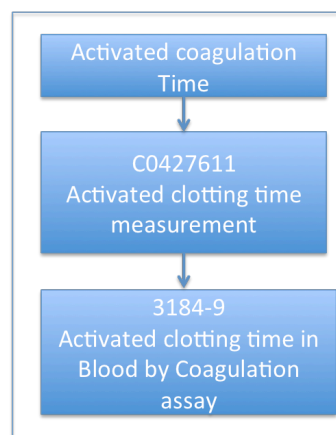


Fig. 4. Example of Lab test Normalization



REFERENCES

- [1] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* 2004;**32**(suppl 1):D267-D70.
- [2] Aronson AR, Lang F-M. An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association* 2010;**17**(3):229-36.
- [3] World Health O. International classification of impairments, disabilities, and handicaps: a manual of classification relating to the consequences of disease, published in accordance with resolution WHA29. 35 of the Twenty-ninth World Health Assembly, May 1976. 1980.
- [4] McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clinical chemistry* 2003;**49**(4):624-33.
- [5] Khan AN, Griffith SP, Moore C, Russell D, Rosario AC, Bertolli J. Standardizing laboratory data by mapping to LOINC. *Journal of the American Medical Informatics Association* 2006;**13**(3):353-55.
- [6] http://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd9cm_reimburse.html.
- [7] <http://loinc.org/downloads/relma>.

Ontology-based Normalization for Disease-Lab Test Relation Extraction

Yaoyun Zhang, PhD, Jingqi Wang, MS, Cui Tao, PhD, Hua Xu, PhD

The School of Biomedical Informatics | The University of Texas Health Science Center at Houston

Introduction

Disease-lab test relation extraction plays an important role in various medical applications such as clinical decision-support systems and phenotype information extraction. However, mentions of diseases and labtests in text have diverse non-standard variations. Those variants need to be normalized into standard codes first to facilitate more universal computational applications. This poster describes our preliminary work on ontology-based normalization of diseases and lab tests, as a fundamental step toward disease-lab test relation extraction.

Ontology Overview

- **UMLS [1]:** Re-organized many controlled vocabularies in the biomedical sciences; a comprehensive ontology of biomedical concepts. E.g., “Diabetes Mellitus” and “dm” are synonyms of the same concept .
- **ICD-9CM [2]:** A hierarchical system of diagnostic codes for classifying diseases. E.g., “(050.0) Variola major” is a subcategory of “(050) Smallpox”.
- **LOINC [3]:** The only publicly available universal standard for laboratory test codes and names. E.g., the test of “blood culture” is under the general category “Microbiology” .

Method

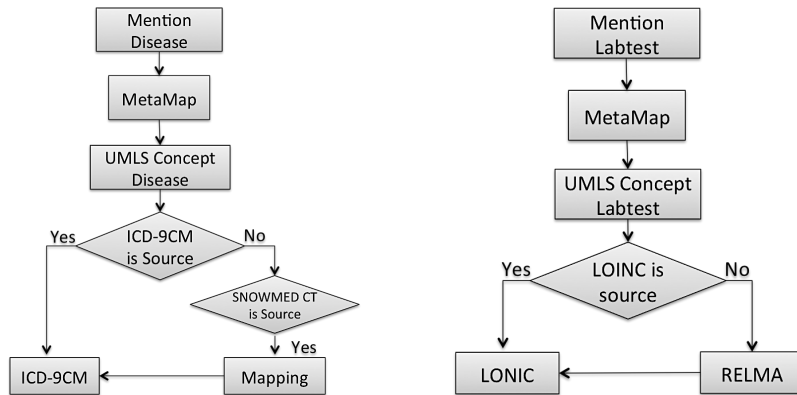


Figure 1 Diagrams of Disease and Lab Test Normalization

Results

- **Precision:** 1) General concepts of diseases/lab tests not so valuable practically. E.g., in relations between “Heart Diseases” and lab tests, “Heart Diseases” include “coronary artery disease”, “arrhythmias” and “congenital heart defects”, etc. 2) Fail to normalize to LOINC by RELMA. E.g., “acanthocyte count” -> 565-2:COLONY [COUNT]:NUM:PT:XXX:ORD:VC.
- **Coverage:** Fail to recognize variants of diseases/lab tests. E.g., “blood film” refers to “blood smear” .

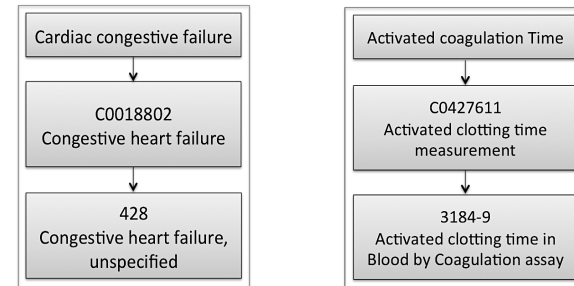


Figure 2 Examples of Disease and Lab Test Normalization

Conclusion

This poster presents the preliminary results of our ontology-based normalization of diseases and lab tests. In the next stage, machine learning methods will be employed for disease and lab test recognition. General concepts of diseases and lab tests need to be filtered. The precision of lab test normalization to LOINC also need to be further improved.

References

1. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research 2004;32(suppl 1):D267-D70.
2. National Center for Health, S. ICD-9-CM: International Classification of Diseases 9th Revision Clinical Modification. US Department of Health and Human Services, Public Health Service, Health Care Financing Administration 2008.
3. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clinical chemistry 2003;49(4):624-33.

Please contact the corresponding author via email: Hua.Xu@uth.tmc.edu