MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)

doi:10.3233/978-1-61499-830-3-511

CrowdMapping: A Crowdsourcing-Based Terminology Mapping Method for Medical Data Standardization

Huajian Mao, Chenyang Chi, Boyu Huang, Haibin Meng, Jinghui Yu, Dongsheng Zhao

Institute of Health Service and Medical Informatics, Academy of Military Medical Sciences of Chinese PLA, Beijing, China

Abstract

Standardized terminology is the prerequisite of data exchange in analysis of clinical processes. However, data from different electronic health record systems are based on idiosyncratic terminology systems, especially when the data is from different hospitals and healthcare organizations. Terminology standardization is necessary for the medical data analysis. We propose a crowdsourcing-based terminology mapping method, CrowdMapping, to standardize the terminology in medical data. CrowdMapping uses a confidential model to determine how terminologies are mapped to a standard system, like ICD-10. The model uses mappings from different health care organizations and evaluates the diversity of the mapping to determine a more sophisticated mapping rule. Further, the CrowdMapping model enables users to rate the mapping result and interact with the model evaluation. CrowdMapping is a work-in-progress system, we present initial results mapping terminologies.

Keywords:

Crowdsourcing; Medical Informatics; Vocabulary, Controlled

Introduction

Analyzing clinical data across medical centers requires mapping local and idiosyncratic information into standardized terminologies. Without this step, clinical data cannot be exchanged, shared, integrated or used in a meaningful way. There are lots of standard terminology systems for representing clinical data, including Logical Observation Identifiers Names and Codes (LOINC) [1], Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) [2] and others, which are now used as international standards. However, in lots of the healthcare organizations, they still use local and idiosyncratic dictionaries, which always reduces the ability to integrate data from multiple organizations. Therefore, in order to understand generated medical information in digital systems, healthcare organizations must often translate the local data into standardized terminologies. The process of translating the local data into standardized terminologies is called mapping. A mapping represents a set of terminology mediation strategies used by clinical and public health organizations to enable health informatione exchange (HIE) within and among health enterprises [3].

At the Institution of Health Service and Medical Informatics of Academy of Military Medical Sciences (AMMS), we are undertaking this mapping process because our medical data center is involved in the creation of a nation-wide health information exchange that requires sharing clinical data across more than 200 first-class hospitals at Grade 3 in China. In the hospital information systems of these 200 hospitals, they use hundreds of dictionaries (e.g., diagnosis, operation, and other terminology dictionaries). Meanwhile, a large part of the terminologies' dictionaries contain a huge amount of entities. How to standardize the terminologies from these 200 hospitals are the most challenging work when integrating their data.

Although there are deviations in terminology standards for different hospitals, there are a number of terms that are generally accepted. In this way, although different local terminology systems are used in different hospitals, a large part of the terminologies are the same. There would be a lot of duplicate work on terminology mapping if each of these 200 hospitals respectively standardizes its terminology system into a specified one.

Besides, if the mapping work is done separately by each of the hospitals, it would probably generate different mappings for some selected terminology. For example, hospital A might map LV_a to ST_a terminology in the standard and hospital B might maps LV_a to a different ST_b terminology. How to eliminate the error mapping for the 200 hospitals and ensure that the local vocabularies in the different hospitals are correctly mapped to the same standard terminologies is challenging work. It would significantly improve the precision of mapping if we had some mechanism to verify the mapping and revise the error mapped ones.

Meanwhile, as there are a huge number of terminology entities in the mappings, it would be also a huge task to map all of them to standard terminologies. However, in practice, only a small part of the local items would be used in the actual healthcare systems. It would improve the efficiency if we considered mapping the terminologies based on their usage frequency.

In this paper, we propose a crowdsourcing based terminology mapping method, CrowdMapping, to help healthcare providers map their local and idiosyncratic data to standard terminologies. This would provide a better use of available medical data for analysis, integration and so on. First, CrowdMapping allows users from different organizations to specify mapping candidates in a standard terminology system like ICD-10. The model uses the mappings from the users and evaluates the diversity of the mapping to determine a most-selected, final best mapping rule if the diversity is less than a threshold value. Otherwise, if the diversity is too large, then some new mappings for the local terminology are generated from some more sophisticated experts. Further, the CrowdMapping model also enables users to give negative ratings to the final selected mapping rule if that mapping would be likely be wrong. CrowdMapping will recalculate the score for that terminology by subtracting a value which is related to the user's reputation. Then, the final best mapping rule is recalculated. CrowdMapping is a work-in-progress system. It will be made public when the system is stable enough for usage.

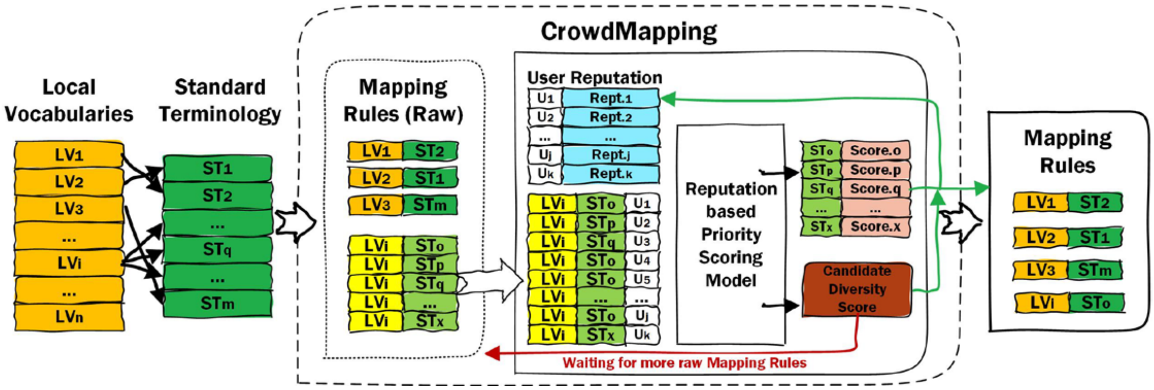

Figure 1 – CrowdMapping Architecture

Methods

Traditionally, standardizing local vocabularies is done by assigning the most likely terminology in the standard system by users. For example, as shown in Figure 1, a user can map LV_1 in the local vocabulary to ST_2 in the standard terminology, same with LV_2 and LV_3 . However, mapping the local vocabulary to a standard terminology is not a trivial work. For example, different users might map some local vocabulary (LV_i in the figure) to different standard terminologies (for example, mapping LV_i to ST_o ,ST_p ,ST_q ,...,ST_x and so on). How to select the most likely candidate as the standard terminology for LV_i would become a problem in this scenario. CrowdMapping proposes a crowdsourcing based solution to solve this problem.

Overview

The target of CrowdMapping is to provide healthcare providers with a platform to map their local vocabularies, stored in dictionaries, to standard terminlogies with high accuracy while keeping the mapping process simple and maintaining high efficiency. There are two basic design ideas in CrowdMapping.

1. CrowdMapping allows different experts to map a local vocabulary to several different candidates. It then uses a rating model to calculate a diversity value for the candidates. If the diversity value is less than a given threshold, a top-rated candidate will be automatically selected as the final terminology for the local vocabulary.
2. The volume of local data to be mapped is always very large, but fortunately, not all of the vocabularies in the local vocabulary systems are used; the usage of the local vocabularies are always biased. Mapping a small part of the local vocabularies would likely standardize most of the local data. With this evidence, CrowdMapping determines the next mapping vocabulary with terminology usage frequency in the local dataset.

Figure 1 shows the design of CrowdMapping. CrowdMapping system consists of three major parts, which are local vocabularies and standard terminology system repositories; a user reputation based priority scoring model; and a final mapping rule repository. The local vocabulary repository contains all the vocabularies which are local and are going to be mapped, while the standard terminology repository is the target to map the local ones to. The standard terminology systems are usually standard terminology systems like ICD-10, SNOMED and so on.

The users map the local vocabulary to standard terminology with their effort to make their decision of finding the most likely one in their opinion for the local one. CrowdMapping then records the user decisions (the raw mapping rule with the tag of which user made the mapping rule in the figure), and passes the raw user mapping rules to a reputation based priority scoring model with the user reputation values. The scoring model will calculate a candidate diversity score for each of local vocabularies (like LV_i). Besides, an ordered candidates queue, the priority scores are calculated by the scoring model. After the calculation, if the diversity is smaller than a given threshold a top-priority standard terminology will be selected as the final mapping value for the local vocabulary. For example, as shown in the figure, if the diversity is small, and ST_o has the highest score, then ST_o will be selected as the mapping value of LV_i . However, when the diversity value is larger than the given threshold, the decisions from different users are not convergent signifying different opinions for the mapping of that local vocabulary. So more raw mapping rules are needed to make the final mapping decision. CrowdMapping will recalculate the diversity and priority scores when a new mapping rule is made for that local vocabulary until the diversity is small enough. After this progress, CrowdMapping will generate a final mapping rules repository for user.

Reputation Based Priority Scoring

In order to describe the reputation based priority scoring model, we would like to define several terms.

User Reputation

User reputation is a value describing the expert level of a user on mapping the vocabularies. If a user always maps local vocabularies to correct standard terminologies, then he will get a higher reputation. Suppose user A has worked on the CrowdMapping platform for some time, and he/she has mapped a set of local terminologies to the standard terminologies. Let's suppose the mapping history of user A is:

`val mappings = Set`{m_1, m_2, …, m_n}

In which, user A has mapped k local vocabularies to standard terminologies correctly, and $n-k$ to error ones. Besides, as we will discuss in the *Mapping Rule Revision* section, the mapping rules can be `'+1'` or `'-1'`. Let's suppose the support number of mapping m_i is p_i , and the negative support number of mapping m_i is n_i. Then, if the mapping m_i is selected as a correct mapping, all the p_i users who have `'+1'`ed this mapping will processed as having mapped correctly one more time, while the n_i error mapped mapping number of the `'-1'`ed users for the mapping m_i will be increased.

Now let's suppose that user A has mapped k local vocabularies correctly, and $n-k$ wrong, and has $'+1'$ ed pc correctly, and pw wrong, and has $'-1'$ ed nc correctly and nw wrong. Then we define the reputation r_i of user A as:

$$r_i = \alpha \frac{k}{k+n-k} + \beta \frac{pc}{pc+pw} + (1-\alpha-\beta)\frac{nc}{nc+nw}$$

where α and β are two parameters which are configured by the CrowdMapping system.

Mapping Set Diversity

Mapping set diversity describes a status of the current mappings for a specific vocabulary. If there are many different mappings for a local vocabulary, that means users having different opinions on the mappings for that local vocabulary will have a high diversity. Otherwise, if all the mappings from different users are the same, then the diversity of the mapping set is 0. Diversity has been used in various areas [4-6]. In the system implementation, Simpson's Diversity Index (SDI) [4] is used to calculate the diversity of the mapping set.

Suppose that the local vocabulary LV_i is mapped to k potential standard terminologies which are $\{ST_i^1, ST_i^2, \dots, ST_i^k\}$, and have n_i^1 users mapping LV_i to ST_i^1, n_i^2 users mapping LV_i to ST_i^2, and n_i^k users mapping LV_i to ST_i^k. Then according to the algorithm of Simpson's Diversity Index, we get

$$D_i = 1 - \frac{n_i^1(n_i^1-1) + n_i^2(n_i^2-1) + \cdots + n_i^k(n_i^k-1)}{N(N-1)}$$

where $N = n_i^1 + n_i^2 + \cdots + n_i^k$.

With this algorithm, the diversity value of the mapping set for LV_i is mapped into a range of [0, 1]. For example, if all the users map the LV_i to ST_i^1, then $N = n_i^1$, so $D_i = 1 - \frac{n_i^1(n_i^1-1)}{N(N-1)} = 0$. Or, if all the users map the LV_i to different standard terminologies, then all the n_i^js equals to 1, which will lead to $D_i = 1 - \frac{\sum n_i^j(1-1)}{N(N-1)} = 1 - 0 = 1$.

Mapping Confidence

Mapping confidence defines the level of how much of the mapping can be trusted. Mapping confidence is always calculated on the selected final mapping rule which has the highest score. It is affected by the mapping count, user revision score, and the mapping diversity. A higher mapping confidence means a higher probability that the mapping can be trusted.

Let's continue with the example used in *Mapping Set Diversity*. Suppose the mapping M_i^j from LV_i to ST_i^j is the mapping which has the highest score. Suppose that the count of the users who map LV_i to ST_i^j is c_i^j, and the diversity of LV_i is D_i, and pc_i^j users $'+1'$ the mapping M_i^j, and nc_i^j users $'-1'$ the mapping M_i^j. The confidence of the mapping M_i^j can be calculated as:

$$C_i = \frac{pc_i^j}{pc_i^j + nc_i^j} * D_i * \frac{count_i^j}{count_{MIN}}$$

Only the mapping with a high confidence larger than a threshold C_{MIN}will be selected as a final mapping for the local vocabulary and be put into the mapping rule repository.

Score Calculation

Based on the definitions above, suppose that n users (U_1, U_2, ..., U_n) are working on mapping the local vocabularies LV_i to standard terminologies. Let's annotate the reputation of U_i as r_i. The user-generated k mappings for LV_i, $\{ST_i^1, ST_i^2, \dots, ST_i^k\}$ (where $k \leq$n). Suppose that those who have mapped the LV_i to ST_i^j are the users (U_a, U_b, … U_x). Then we will calculate a temporal reputation summary of the mapping M_i^j from LV_i to ST_i^j:

$$Reput_j = \sum_{i=a}^{x} r_i$$

and then the score for the mapping M_i^j is calculated as:

$$Score_j = \frac{Reput_j}{\sum_{i=1}^{k} Reput_i}$$

The reputation based priority scoring model will calculate all the mappings and their scores and generate an scored terminology map as $\{ST_o{:}\,Score_o$, $ST_p{:}\,Score_p$, $ST_q{:}\,Score_q$, …, $ST_x{:}\,Score_x\}$.

After the scores are computed, the CrowdMapping method will calculate the mapping confidence value for the mapping with the highest score. CrowdMapping will compare it with a given threshold value, if the diversity score is less than the threshold, CrowdMapping will select the standard terminology $ST_o$with the highest score (for example, $Score_o$) as the mapping to terminology for the local vocabulary (for example, LV_i in the figure). CrowdMapping will also update the user reputation table according to the last mapping decisions. In this way, the users who have mapped LV_i to ST_o will get a higher reputation value, while the users mapped LV_i to other standard terminology will be reduced on their reputation. However, if the confidence value of the mapping is too small, suggesitng that the users do not have a consistent decision on how to map LV_i to the standard terminology, CrowdMapping will not do the following steps as described above. It will wait for more user raw mapping rules.

Algorithm 1 shows the process of reputation based priority scoring method of finding the most likely mapping M_i^j for LV_i. The CrowdMapping system will use this algorithm to process the mappings for all the local vocabularies.

With these processes, a final mapping decision will be made based on all the user mapping decisions, as shown in the Mapping Rules in the figure.

Algorithm 1: reputation based score algorithm for $\boldsymbol{LV_i}$

```
mappings ← all the mappings of LVi with user info

Di ← calculate diversity of the mappings
if Di > Dthreshold, then
    users ← get users of the mappings
    for each user in users, do
        ri ← calculate user reputation for each user
    end
    scoredMap ← init scored terminology mappings
    for each mapping in mappings, do
        Scorei ← calculate Score of mapping
        Ci ← calculate confidence of mapping
        if Ci > CMIN, then
            Add (STi, Scorei) to scoredMap
        end
    end

    orderedScoredMap ← sort scoredMap by score

    return top of orderedScoredMap
else
    return NULL
end
```

Frequency Based Local Vocabulary Candidate

Local vocabularies are always having a large volume of entries. In practice, only a limited part of the local vocabularies are used in the healthcare systems. So different entries in the vocabularies have different significance for health care systems. It would significantly improve the work efficiency if the most 'important' entries in the local vocabulary are mapped first. So CrowdMapping defines a usage frequency based priority value for local entry candidate selection. If an entry in the vocabulary is used quite frequently in a health care setting it will have a higher priority for mapping.

However, the frequency is calculated from online health care systems. In some case, users might not be able to get the frequency. In these situations CrowdMapping will use a similar frequency statistic from another health care organization. If no organization has the frequency information, then equal priority will be assigned to the local vocabulary entries. In this way, the next local vocabulary entry to be processed is selected randomly.

Mapping Rule Revision

After the steps described above are performed a usable mapping rules repository will be generated. And users from different healthcare organizations can use these mapping rules to automatically standardize their local vocabularies. However, CrowdMapping is a machine-based automation tool, so there might exist erroroneous mappings in the system-generated rules. An official selected mapping may also have expired and/or become invalid. For these reasons, CrowdMapping has a mapping rule revision module. In this module, the users can `'+1'` a mapping rule or `'-1'` it. And these actions will trigger the mapping confidence computation. If there are too many users `'-1'`ing that mapping rule, the mapping confidence will decrease. When the confidence is lower than a threshold, the mapping rule will be removed from the final mapping rule repository to the raw mapping rule repository to await more user mapping decision.

Results

As CrowdMapping is still a work-in-progress project, we have not performed large scale testing. However, with using the key design idea in a demo platform, we found that CrowdMapping combines the mapping decisions from multiple users and selects the most likely rule for the local vocabularies which makes the rule mapping decision more precise.

When the platform is ready, we will evaluate CrowdMapping with the traditional mapping method on mapping precision and the processing time. Also, we will make CrowdMapping as a public web application. All the users from healthcare organizations can use our service together with the effort from the experts all over the world.

Discussion

Terminology mapping is important work for the medical informatics community. There is already a large body of work describing various aspects of mapping local codes to standard terminology systems from both academic and industry researchers.

Several studies [7-11] have evaluated different automated tools to assist with mapping local vocabularies to other terminology systems. Yet, even with the best available automated tools, expert human review is still needed to resolve computer generated candidate mappings. Also, because local and standard vocabularies evolve, the burden of maintaining the mappings is significant, ongoing, and easily underestimated. Therefore, all healthcare organizations whether data senders, receivers, or both, require people, processes, and tools to support mapping activities.

It is a complex and resource intensive job to map local terms to standard terminology systems. Even a sophisticated person with a good understanding of the corresponding terminology system, he/she might lack the specific knowledge required to correctly map all of their local terminologies [3,12,13]. How to efficiently combine the effort from the experts of the institutions to improve the mapping precision is a necessary work. Crowdsourcing has been used in many fields to improve the work efficiency. In the terminology mapping, several work [3,12,13] have been done to make mapping more efficient and effective with crouwdsourcing. However, they are strongly integrated with special standardized terminlogies, for example, LOINC.

CrowdMapping allows users from different organizations to specify mapping candidates in standard terminology systems like ICD-10. CrowdMapping uses the mappings from the users and evaluates the diversity of the mapping to determine a most-frequently selected final best mapping rule. This strategy would likely improve the accuracy of the mapping. Mapping set diversity and the mapping confidence model are used in CrowdMapping to automatically divide the 'hard' mapping tasks from the 'easy' ones (a local entry with a high mapping diversity is always a 'hard' mapping task). CrowdMapping can let the 'right' person to do the 'right' thing; the 'hard' mapping tasks are assigned to sophisticated experts to make the mapping more precise.

Conclusion

In this paper, we propose CrowdMapping, a platform for generating terminology mapping rules that leverages the crowd effort. CrowdMapping uses an algorithm which considers both the user reputation and crowd selection, which likely leads to a more confident mapping result. Using the demo program, we found that CrowdMapping reduces the mapping time and increases precision. Terminology mapping is a very common step for medical data analysis, we will deploy an online service for public use as soon as the platform is more stable.

Acknowledgements

The authors would like to thank the anonymous reviewers for their insight, comments, and kindsuggestions. Their valuable comments helped improve this paper a lot.

References

[1] R.A. Cote, College of American Pathologists, *Systematized Nomenclature of Medicine*, College of American Pathologists, 1977.

[2] B.E. Dixon, J. Hook, D.J. Veerman, Learning from the crowd in terminology mapping: the loinc experience, *Lab Med* **46** (2015), 168-174.

[3] C. McDonald, S. Huff, J. Suico, K. Mercer, Logical observation identifiers names and codes (loinc R) users' guide, in, Regenstrief Institute, Indianapolis, 2004.

[4] E.H. Simpson, Measurement of Diversity, *Nature* **163** (1949), 688.

[5] M. Nei, Analysis of gene diversity in subdivided populations, in: *Proceedings of the National Academy of Sciences*, 1973, pp. 3321-3323.

[6] A.E. Magurran, *Why diversity? In ecological diversity and its measurement*, 1988.

[7] A.N. Khan, S.P. Griffith, C. Moore, D. Russell, A.C. Rosario, J. Bertolli, Standardizing labora- tory data by mapping to loinc., *J Am Med Inform Assoc* **13** (2006), 353-355.

[8] L.M. Lau, K. Johnson, K. Monson, S.H. Lam, S.M. Huff, A method for the automated mapping of laboratory results to loinc. , in: *Proceedings of the AMIA Symposium*, 2000.

[9] J.Y. Sun, Y. Sun, A system for automated lexical mapping., *J Am Med*

Inform Assoc **7** (2006), 334-343.
[10] K.A. Zollo, S.M. Huff, Automated mapping of observation codes using extensional definitions, *J Am Med Inform Assoc* **7** (2000), 586-592.
[11] C. Zunner, T. Burkle, H.U. Prokosch, T. Ganslandt, Mapping local laboratory interface terms to loinc at a german university hospital using relma v. 5: a semi-automated approach, *J Am Med Inform Assoc* **20** (2013), 293-297.
[12] M.C. Lin, D.J. Vreeman, C. McDonald, S.M. Huff, A characterization of local loinc mapping for la- boratory tests in three large institutions, *Method Inform Med* **50** (2011), 105.
[13] D.M. Baorto, J.J. Cimino, C.A. Parvin, M.G. Kahn, Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC), *Int J Med Inform* **51** (1998), 29-37.

Address for correspondence

Dongsheng Zhao, Mail: dszhao@bmi.ac.cn