Automatic coding of Free-Text Medication Data recorded by Research Coordinators

Laritza M. Rodriguez, MD, PhD, Vojtech Huser, MD, PhD, Olivier Bodenreider, MD, PhD, James J. Cimino, MD, PhD
Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health Clinical Center, National Institutes of Health, Bethesda, MD

Abstract

Medication reconciliation and clinician-friendly recording of medication is an ongoing informatics challenge. Users often prefer unrestricted free-text entry. We sought to see if simple Natural Language Processing (NLP) methods can be used to convert real-life medication records into coded concepts without additional interface changes to clinicians. We used data from the National Institutes of Health Clinical Center (NIH CC). Outpatient medications are recorded by research nurses using a semi-structured form that does not alert for incorrect data entry. We present data on accuracy of the NLP pipeline and suggest improvements to RxNorm tools.

Objective: To utilize simplified automatic NLP tools for drug entry harmonization.
Materials and Methods: We used National Library of Medicine's RxNorm API through the RxMix tool (http://rxnav.nlm.nih.gov/, accessed on 2014-01-30) to convert free-text medication entries into RxNorm concepts, namely RxNorm ingredients**.** Medication free text entries were first preprocessed with regular expressions to extract the drug names, drug strength, and drug forms from the original string that also included dosing and frequency information. The resulting simplified string was processed using the function *getApproxMatch* to find matching concepts in RxNorm (RxCUIs). The RxCUI was then processed using the *getAllRelatedInfo* function to obtain the ingredient and additional information. We used the concept match score and rank provided by the *getApproxMatch* function to further consider only well-matched concepts. For example, from the original input string: "*Metronidazole 250 mg , 2 tablet , PO , Every 12 hours.*", we first extract the drug entity "*metronidazole 250 mg*", which maps to the RxNorm concept *"Metronidazole 250 MG"* (316300) with a score of 100. Its ingredient is *"Metronidazole"* (6922). The remainder of the input string drug dose and frequency can be used to estimate drug doses or as part of a free text drug entry system automatic fill option.
Results: The original 43,303 medication strings were preprocessed into distinct 9,466 drug input strings. A total of 5,680 (60%) strings mapped to either RxNorm Ingredient (IN) concept or Clinical Drug (SCD) or both. Online appendix at http://dx.doi.org/10.6084/m9.figshare.960060 shows complete numerical results together with the impact on the string mapping counts when score thresholds are used to classify strings into two scenarios: (A) likely correction to a single target concept (score 80-100; single misspelled character: *"atenlol"* to *"atenolol"*) or (B) nurse may be asked to pick the best term from several matching contexts (score 50-79; multiple misspellings). A total of 502 concepts had either no matches or a score of less than 50 most of which are not ingredients or meaningless strings eg: *"vitamin juice", "zyrec"*.
Evaluation: The evaluation of the automatic mapping was done by one of the authors (clinician LMR). The unmatched set of strings was analyzed to detect those that should have found a match and did not. Two mapping errors were found: *"codeine"* and *"chlorpheniramine"* mapped to two unrelated ingredient names with the highest score of 100. (These errors were reported to and corrected by the RxMix team). Overall, the method had a 0.99 precision, 0.97 recall and 0.98 F1-measure. The manual evaluation was performed by only one clinician.
Conclusions: The results demonstrate that simple NLP methods that use the existing free RxNorm API can be used to convert free text medications records into RxNorm concepts. Coded concepts offer researchers better data queries that can utilize drug class hierarchies and outperform the existing free-text search. The quality of the mapping using the automated string matching is excellent provided the total number of input strings with only two mapping errors (employing also the above score filters). Although the approximate matching function of RxMix was primarily developed to support matching of clinical drug names, not ingredients alone, we found it useful in the context of this experiment where our input data had various degrees of specification.

References

1.   Zhou L, Plasek JM, Mahoney LM, Karipineni N, Chang F, Yan X, et al. Using Medical Text Extraction, Reasoning and Mapping System (MTERMS) to process medication information in outpatient clinical notes. AMIA Annu Symp Proc AMIA Symp AMIA Symp. 2011;2011:1639–48.