

Extraction of Drug-Drug Interactions from Drug Product Labeling Tables

Nikola Milošević, MSc¹, Anay Gupta², Austin Chen², Steven T. DeMarco², Joshua Le², PhD², Jodi Schneider³, Yifan Ning², Goran Nenadić, PhD¹, Richard D. Boyce²

¹University of Manchester, Manchester, UK;

²University of Pittsburgh, Pittsburgh, PA, USA;

³University of Illinois Urbana-Champaign, IL, USA;

Introduction

Many thousands of patients are harmed each year as a result of improper management of interacting prescription drug combinations¹. In the United States, Structured Product Labels (SPLs) are mandated to provide information about potential drug-drug interactions. Such information is provided in either unstructured text or in diversely formatted XML tables. For example, at the time of writing, more than 1200 structured product labels representing more than 870 clinical drugs contain table-formatted PDDI information. While tables present large amount of multidimensional information in a visually condensed format, they are not designed for computational processing and understanding. This presents a barrier to systems designed to help drug experts search, retrieve, and synthesize this information for clinical decision support.

Methodology

We present a method to extract PDDI mentions from tables. It consists of five of steps: (1) table detection, (2) functional table analysis, (3) structural table analysis, (4) semantic annotation, and (5) information extraction. Table detection in XML documents is trivial, consisting on finding table tags. The functional analysis uses a set of heuristics about cell's content, position and its neighboring cells to determine cell's function in the table. Cells can have navigational function (headers, super-rows, stubs) or data presentation function (data cells). The structural table analysis uses the heuristics about cell's function, position and alignment to disentangle the relationships between navigational and data cells. In other words, it finds cells that describe presented data. The functional and structural analyses were performed by TableAnnotator², which was modified to allow processing of SPLs. In step (4), we have manually identified words mentioned in table column headings and mapped them to eight key categories of information (e.g., drug class or name, effect on drug, and recommendation or comment). We then annotate named entities present in table cells using MetaMap and UMLS semantic types. In the final step, we use rules and heuristics to extract PDDI mentions and related information from the tables. Each heuristic is being tested for recall and precision. For example, we first attempt to infer an interaction between a label drug and all named entities mentioned within in a column having a 'drug class or name' header. The results are then filtered to include only those entities assigned a UMLS semantic type such as 'clinical drug' and 'pharmacological substance'. Further filtering involves using the ATC (Anatomical Therapeutic Chemical) terminology to keep only drug ingredient mentions.

Results

We were able to group 340 table column headers into 8 categories of information. A pilot evaluation showed that the proposed approach has high precision (~95%) but relatively weak recall (65%-75%) for extracting drug-drug interaction pairs. The future work will expand the methodology to involve additional pharmacological information and several new heuristics.

Conclusion

This is the first attempt to develop and test a method for automatically extracting PDDI mentions from structured product labeling. We envisage that the methodology will facilitate PDDI knowledge base development and make it easier for regulators to monitor for PDDI information changes in DailyMed labels.

References

1. Magro L, Moretti U, Leone R. Epidemiology and characteristics of adverse drug reactions caused by drug-drug interactions. *Expert Opin Drug Saf.* 2012 Jan;11(1):83-94.
2. Milosevic, N., Gregson, C., Hernandez, R. and Nenadic, G. Disentangling the Structure of Tables in Scientific Literature. In *International Conference on Applications of Natural Language to Information Systems.* June 2016: 162-174.