Relation Extraction from Biomedical Literature on Pharmacokinetic Natural Product-Drug Interactions

Sanya B. Taneja[1], Patrick W. Ndungu[1], Mary F. Paine[2], Sandra L. Kane-Gill[1], Richard D. Boyce[1]

[1]University of Pittsburgh, Pittsburgh, PA, [2]Washington State University, Spokane, Washington

Introduction

Co-consumption of botanical and other natural products (NPs) with pharmaceutical drugs can lead to pharmacokinetic NP-drug interactions (NPDIs), which can further lead to unwanted drug response[1]. Computational approaches such as semantic relation extraction from *in vitro* and clinical pharmacokinetic studies related to the topic can provide insights about the mechanisms of these NPDIs. We extracted predications (subject-predicate-object triples) from full text articles focused on pharmacokinetic drug interactions involving the model NP green tea using two relation extraction systems, related the information to existing biomedical resources, and compared the results to human data extraction.

Methods: Full texts of 13 PubMed-indexed articles related to pharmacokinetic green tea-drug interactions were obtained from PubMed Central and PDF files. SemRep (v1.8)[2] and the Integrated Network and Dynamical Reasoning Assembler (INDRA)[3] with the REACH[4] biological reader were used to extract predications from the full texts. All subjects and objects in the predications were mapped to the Unified Medical Language System (UMLS) concepts using MetaMap. The predications were compared with data from the Center of Excellence for Natural Product Drug Interaction Research (NaPDI Center) public repository[1]. Data extracted in the repository by humans from the same articles was translated to predications using relations from the Open Biological and Biomedical Ontology (OBO) Foundry ontologies and established as the ground truth.

Results

Counts of pharmacokinetic-related predicates from the NaPDI Center repository, SemRep, and INDRA/REACH are shown. Translation of the repository data resulted in 179 predications. 522 predications were extracted from SemRep. 123 predications were extracted from INDRA/REACH, of which 94 were relevant to pharmacokinetic NPDIs (76.4%); 65 were green tea specific (52.8%). 193 predications from SemRep were related to pharmacokinetic interactions, 128 were relevant to pharmacokinetic NPDIs (66.3%), and 77 (39.9%) were green tea specific. The recall values for SemRep and INDRA/REACH were 0.31 and 0.20, respectively.

Predicate	NaPDI Center Repository	SemRep	INDRA/ REACH
INHIBITS	80	38	79
INTERACTS WITH	46	54	0
ACTIVATES/STIMULATES	0	13	40
SUBSTRATE OF*	23	0	0
OTHER (PHARMACOKINETIC)	30	88	4

Conclusion: We present a baseline evaluation of relation extraction from full texts of a subset of articles focused on green tea-related pharmacokinetic interactions. The results are analogous to prior studies, however, certain predicate types such as 'substrate of' were not captured by either of the systems as they are not included in the relations extracted by the systems. It may be possible to map these predicates to alternative relations. We are currently assessing the correctness of inferences and evaluation beyond the recall for the broader set of articles from our search strategy. Although there are several challenges in extracting relevant information, semantic relation extraction is a scalable approach to find associations between biomedical entities beyond simple named entity recognition to inform scientists of evidence of prior work.

References

1. Birer-Williams C, Gufford BT, Chou E, Alilio M, VanAlstine S, Morley RE, et al. A New Data Repository for Pharmacokinetic Natural Product-Drug Interactions: From Chemical Characterization to Clinical Studies. Drug Metab Dispos. 2020 Oct;48(10):1104–12.
2. Kilicoglu H, Rosemblat G, Fiszman M, Shin D. Broad-coverage biomedical relation extraction with SemRep. BMC Bioinformatics. 2020 May 14;21(1):188.
3. Gyori BM, Bachman JA, Subramanian K, Muhlich JL, Galescu L, Sorger PK. From word models to executable models of signaling networks using automated assembly. Mol Syst Biol. 2017 Nov 1;13(11):954.
4. Valenzuela-Escárcega MA, Babur Ö, Hahn-Powell G, Bell D, Hicks T, Noriega-Atala E, et al. Large-scale automated machine reading discovers new cancer-driving mechanisms. Database J Biol Databases Curation. 2018 01;2018.