Providing Molecular Characterization for Unexplained Adverse Drug Reactions

François-Élie Calvier, Pierre Monnin, Miguel Boland, Patryk Jarnot, Emmanuel Bresso, Malika Smaïl-Tabbone, Adrien Coulet, Cedric Bousquet

Providing Molecular Characterization for Unexplained Adverse Drug Reactions

François-Elie Calvier[a*], Pierre Monnin [b], Miguel Boland [c], Patryk Jarnot [d], Emmanuel Bresso [b], Malika Smail-Tabbone [b], Adrien Coulet [b e], Cedric Bousquet [a f]

[a] SSPIM, Unit of Public Health and Medical Informatics, CHU University Hospital of Saint Etienne, France
[b] Université de Lorraine, CNRS, Inria, LORIA, Nancy, France [c] EMBL-EBI, Hinxton, United Kingdom
*[d] Silesian University of Technology, Gliwice, Poland [e] Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, California [f] Sorbonne Université, Inserm, université Paris 13, LIMICS, Paris, France * francois.calvier@chu-st-etienne.fr*

Keywords:
Machine learning, Adverse drug reaction, Feature selection

Introduction

Mining large drug-oriented knowledge graphs enables predicting Adverse Drug Reactions (ADRs) [1, 2]. Indeed, these graphs encompass knowledge elements about the molecular mechanism of drugs (e.g. drug targets, Gene Ontology annotations, gene variations, pathways). However, only few works [3] explored further these graphs in the search for mechanistic explanation for this type of events. We assume that features documenting molecular mechanisms that take part in the prediction are particularly interesting features, since they may provide novel knowledge for the mechanism that may be underlying an ADR. We propose to explore PGxLOD [4], a knowledge graph built around drugs and pharmacogenomic processes in which they are involved, through the lens of several ADR datasets, each focusing on a particular type of ADRs. Particularly, we propose to use features resulting from the exploration of PGxLOD in a prediction task where best predictive features will be considered as potential elements of explanation.

Methods

Following reference studies, we distinguish drugs associated with three families of ADRs: (1) Human Leukocyte Antigen system related ADR, (*n=104*); (2) Drug Induced Liver Injury (*n=1036*); (3) Severe Cutaneous Adverse Reactions (*n=874*). In each case, we add a set of negative drugs, i.e., drugs most probably not associated with an ADR of the family in question, with distinct methods. Regarding (1), positive drugs are supposed to be associated with type B effects (idiosyncratic), then we selected a set of drugs associated with type A effects (dose-related) to define negative examples. Negative drugs associated with (2) and (3) were provided by reference studies, but were cleaned by litterature review. Drugs from the resulting list are then mapped to those defined in PGxLOD, to enable exploring the graph for features describing each drug. Various graph exploration methods are tested in parallel to generate various sets of features. First, all nodes in a neighbourhood of 5 hops from a drug are considered. Second, we select only nodes more connected to positive drugs than to negative and inversely, using a threshold ratio. Third, we use neighbouring nodes, paths to them and their ontology classes to build a feature matrix.

Results

After setting negative examples and mapping with PGxLOD, drug lists (1) (2) and (3) are respectively composed of 55/452/106 positive and 747/268/334 negative drugs. Results in regards with graph exploration, feature selection and evaluation are ongoing at the time we write this abstract.

Discussion & Conclusion

We hypothesize that PGxLOD encompasses an original set of features potentially intangible in the databases it is connecting. We will present how their iterative heuristical selection permit to identify candidate features to potentially explain ADRs. Groups of ADRs from Side Effect Resource (SIDER) and expert reviewing will be employed to evaluate our method. We initiated this work during the BioHackathon 2018 Paris (https://bh2018paris.info), and are actively pursuing.

Acknowledgements

PractiKPharma project (http://practikpharma.loria.fr/), ANR-15-CE23-0028, funded by the French Research Agency

References

[1] M. Kamdar, M. A. Musen, PhLeGrA: Graph analytic in pharmacology over the web of life sciences linked open data, Int. Conference on World Wide Web **26** (2017), 321-329.
[2] E. Muñoz et al., Facilitating prediction of adverse drug reactions by using knowledge graphs and multi-label learning models, Briefings un bioinformatics **20** (2017), 190-202.
[3] E. Bresso et al., Integrative relational machine learning approach for understanding drug side effect profiles, BMC Bioinformatics **14** (2017), 207.
[4] P. Monnin et al. PGxO and PGxLOD: a reconciliation of pharmacogenomic knowledge of various provenances enabling further comparison, bioRxiv (2018), 390971.