

Using Drug Similarities for Discovery of Possible Adverse Reactions

Emir Muñoz, MEng^{1,2}, Vít Nováček, PhD², Pierre-Yves Vandenbussche, PhD¹

¹Fujitsu Ireland Ltd., Co. Dublin, Ireland;

²Insight Centre for Data Analytics at NUI Galway, Co. Galway, Ireland

Abstract

We propose a new computational method for discovery of possible adverse drug reactions. The method consists of two key steps. First we use openly available resources to semi-automatically compile a consolidated data set describing drugs and their features (e.g., chemical structure, related targets, indications or known adverse reaction). The data set is represented as a graph, which allows for definition of graph-based similarity metrics. The metrics can then be used for propagating known adverse reactions between similar drugs, which leads to weighted (i.e., ranked) predictions of previously unknown links between drugs and their possible side effects. We implemented the proposed method in the form of a software prototype and evaluated our approach by discarding known drug-side effect links from our data and checking whether our prototype is able to re-discover them. As this is an evaluation methodology used by several recent state of the art approaches, we could compare our results with them. Our approach scored best in all widely used metrics like precision, recall or the ratio of relevant predictions present among the top ranked results. The improvement was as much as 125.79% over the next best approach. For instance, the F1 score was 0.5606 (66.35% better than the next best method). Most importantly, in 95.32% of cases, the top five results contain at least one, but typically three correctly predicted side effect (36.05% better than the second best approach).

Introduction

Adverse Drug Reactions (ADRs)^a can severely limit the intended benefit of drugs and accounts for a large number of hospital admissions, 42% of which could be prevented^{1,2}. ADRs can result in reduction of the patients' quality of life or even death in extreme cases³. The use of machine learning techniques has now become a common practice to improve drug safety and in particular to detect ADRs. However, many of the state of the art side effect detection systems and procedures depend on patient records or explicit incident reports³⁻⁵ and therefore assume ADRs already demonstrated within a population.

Stakeholders in the drug development and administration lifecycle could greatly benefit from a technique that would help palliating drug's ADRs before it is released on the market. The presented work addresses the area of computational side effect discovery using information in openly available biomedical databases. In recent years, increasing volume of biomedical data has been openly published online. This includes structured resources like Drugbank⁶ and SIDER⁷ that are represented in a machine-readable and interchangeable standard format in the Bio2RDF project⁸. The uniform format allows for easy combination of these resources that can help to get different viewpoints on biomedical facts within one interlinked resource --- *knowledge graph*, an increasingly used umbrella term for loosely structured graph-based knowledge representation⁹. Knowledge graphs are well suited to discovery of implicit knowledge hidden in the data, which we utilise in our approach to discovery of adverse drug reactions. We infer new links between drugs and side effects using side effect propagation along drug similarity relationships computed using the contents of consolidated Bio2RDF graphs.

The key contributions of our approach can be summarized as follows: (1) Best results in seven metrics traditionally used in the field (e.g., precision, recall or number of correct predictions among the top-ranked results) when compared to recent related works¹⁰⁻¹⁴. (2) Best performance in being able to discover actual side effects and rank them so that they appear at the top of the results. In 95.32% of cases, the top 5 results contain at least one correctly predicted side effect (36.05% better than the second best approach). Moreover, the top 5 results typically contain at least 3 correctly predicted side effects. For drug of certain types (like NSAIDs or barbiturates), all the top 5 results are typically correct discoveries^b. (3) Superior flexibility - our prototype is able to incorporate many relevant data sets automatically, while all other related approaches would need to either develop additional ad hoc pre-processing tools, or replicate the presented technology.

^aNote that we also use several synonyms of ADRs like "adverse drug events" or "drug side effects" interchangeably in the paper.

^bThis is a very important aspect of the presented method, since in many practical applications, it is crucial to provide high-quality results among the top few ones. To give two examples: In clinical applications, physicians have very little time and therefore any automated predictions the use for their decisions have to be concise and highly reliable. In pharmaceutical research, the computational discoveries have to be tested in expensive and long laboratory experiments, and thus a more concise and reliable method like ours can potentially save money and time.

The expected benefits of the presented technology for its target audiences are:

- fully automated prediction of possible side effects;
- flexibility due to support for semi-automated incorporation of new training data sets;
- applicability to decision support in
 - clinical practice -- offering instant, reliable and concise feedback on possible side effects of drugs in daily use, helping to save lives and prevent complications in patients;
 - pharmacological research -- providing comprehensive lists of potential side effects of new compounds based in their similarity to existing drugs, helping to develop drugs faster and save money, while contributing to their safety.

A current limitation of the presented technology is its ability to make predictions about slightly smaller number of drugs than some related approaches. Still, this does not prevent from practical applications bringing the above-mentioned benefits, as explained in detail in the discussion section.

System Overview

The drug-side effect link prediction system consists of two phases: P1) an offline phase where the data processing takes place (integration of the heterogeneous data sources to build a knowledge graph in the RDF format¹⁵, which is then used for computation of drug similarities), and P2) an on-line phase that provides interfaces for users to query the drug similarities database. Each phase consist of different module interactions to either build the similarities database, or query it. Figure 1 shows the system architecture and components involved in both phases.

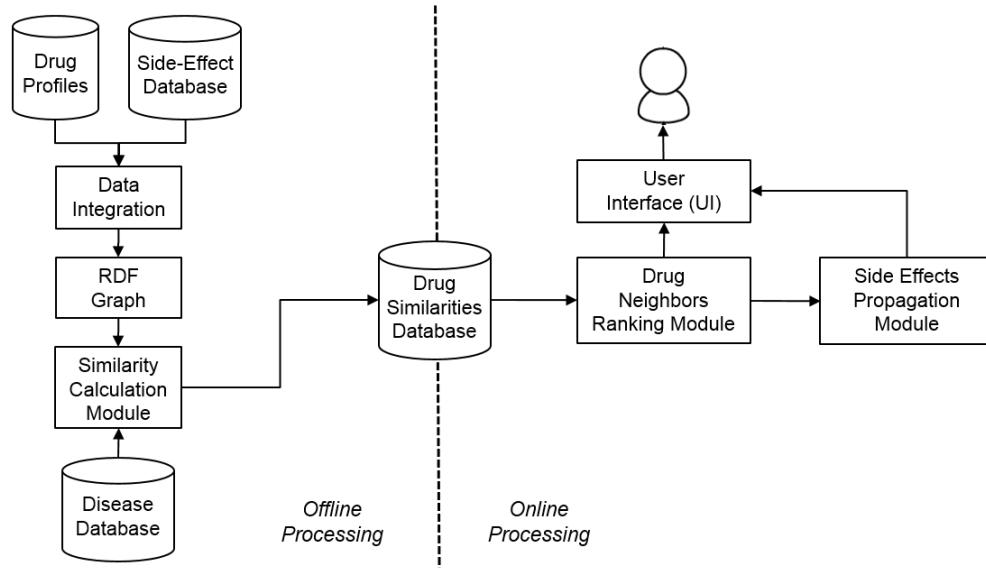


Figure 1. System architecture.

Methods

We gathered all data sets into the pivot RDF format. The formal specification of RDF is convenient, as it allows for the integration of data while preserving their original semantics. We use a graph representation to model the underlying drug-related background knowledge.

Data sources

The main data sets we used are DrugBank for drugs, SIDER for drug side-effects, and PubChem for compound IDs which are used to link drugs in DrugBank to the ones in SIDER. Although all the mentioned sources have public access dumps, our method takes advantage of the graph representation using RDF, therefore we use the transformed data made available in the Bio2RDF project⁸ (release 4, accessed in December, 2015).

Table 1. Data sets used for integration.

Data set	Type	Source and version	Content
DrugBank	Database	Bio2RDF v2015-12-06	Drug types, chemical information
SIDER	Database	Bio2RDF v2015-12-06	Side effects of drugs

The RDF graphs with all data sets are stored in a Apache Jena Fuseki^c triplestore and accessed using the W3C recommendation SPARQL 1.1^d query language for RDF.

Data processing

The first step in the data processing pipeline is indexing the Bio2RDF data sets to facilitate a uniform access to them. As shown in Figure 1, the data sources considered herein are drug profiles and side effects, which are integrated by the module data integration, to later make them available for querying as an RDF graph using HTTP SPARQL requests. Even when we use different public sources to build the RDF knowledge graph, by design our method is able to work with other sources such as post-market reports. Because of the flexibility of RDF, other structured sources can be used with information about drugs (e.g., chemical, biological, phenotypic), side-effects, drug classification, and diseases.

The number of unique approved small-molecule drugs gathered from the structured data sets was 731, while the number of side-effects gathered was 4,652. Finally, the resulting RDF knowledge graph contains a total number of 10.7 million unique statements.

Measures

Resource Features Vector. Given a resource (e.g., a drug node in our specific case), we extract a set of features that represent the connections between the resource and other resources in the graph. Formally, we extract the set of incoming and outgoing relationships of a resource X by using query patterns $(?, ?, X)$ and $(X, ?, ?)$, respectively. For example, let A be the set of features for resource a in Figure 2. Node a in the RDF knowledge graph has outgoing relations to the resources c , e , and f with predicates ℓ_1 , ℓ_3 , and ℓ_4 , respectively; and a unique incoming relation from resource d with predicate ℓ_2 . Then A is defined as follows:

$$A = \text{Features}_{\text{LD}}(a) = \{\{(\ell_1, c), (\ell_3, e), (\ell_4, f)\}, \{(\ell_2, d)\}\}.$$

Similarly, the set of features for resource b is given by the set B .

$$B = \text{Features}_{\text{LD}}(b) = \{\{(\ell_4, e), (\ell_4, f), (\ell_5, g)\}, \{(\ell_2, d)\}\}$$

In our experiments, we manually remove features with functional properties, properties that can have only one (unique) value for each resource, *i.e.*, there are no two distinct resources with the same property value (e.g., identifiers).

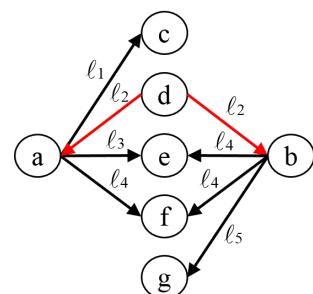


Figure 2. Example RDF graph.

Similarity Metrics. In the present experiment we use the 3w-Jaccard binary similarity measure¹⁶ between two RDF nodes based on their corresponding feature

^c<https://jena.apache.org/>

^d<https://www.w3.org/TR/sparql11-query/>

sets^e. Intuitively, the more features two nodes have in common, the more similar they are. Let x be the size of the set of features shared by \mathbf{a} and \mathbf{b} , y the size of the set of features only present in A , and z the size of the set of features only present in B . The 3w-Jaccard similarity is defined to weight higher common features, and weight lower discriminating features, i.e., those only present in A or B.

$$S_{3W\text{-JACCARD}}(\mathbf{a}, \mathbf{b}) = \frac{3x}{3x + y + z}, \text{ with } 0 \leq S_{3W\text{-JACCARD}}(\mathbf{a}, \mathbf{b}) \leq 1.$$

Example 1 Consider two drugs, namely, Phenacetin (identified by `dBank:DB03783`) and Acetaminophen (identified by `dBank:DB00316`), with a subset of their relations including: label, brand name, dosage form, and target proteins. As Figure 3 shows, these two drugs share some relations with the resources "Humans and other mammals"@en, `dBank:capsuleOral`, `dBank:liquidOral`, `dBank:analgesics,Non-narcotic`, and `dBank:target-20`.

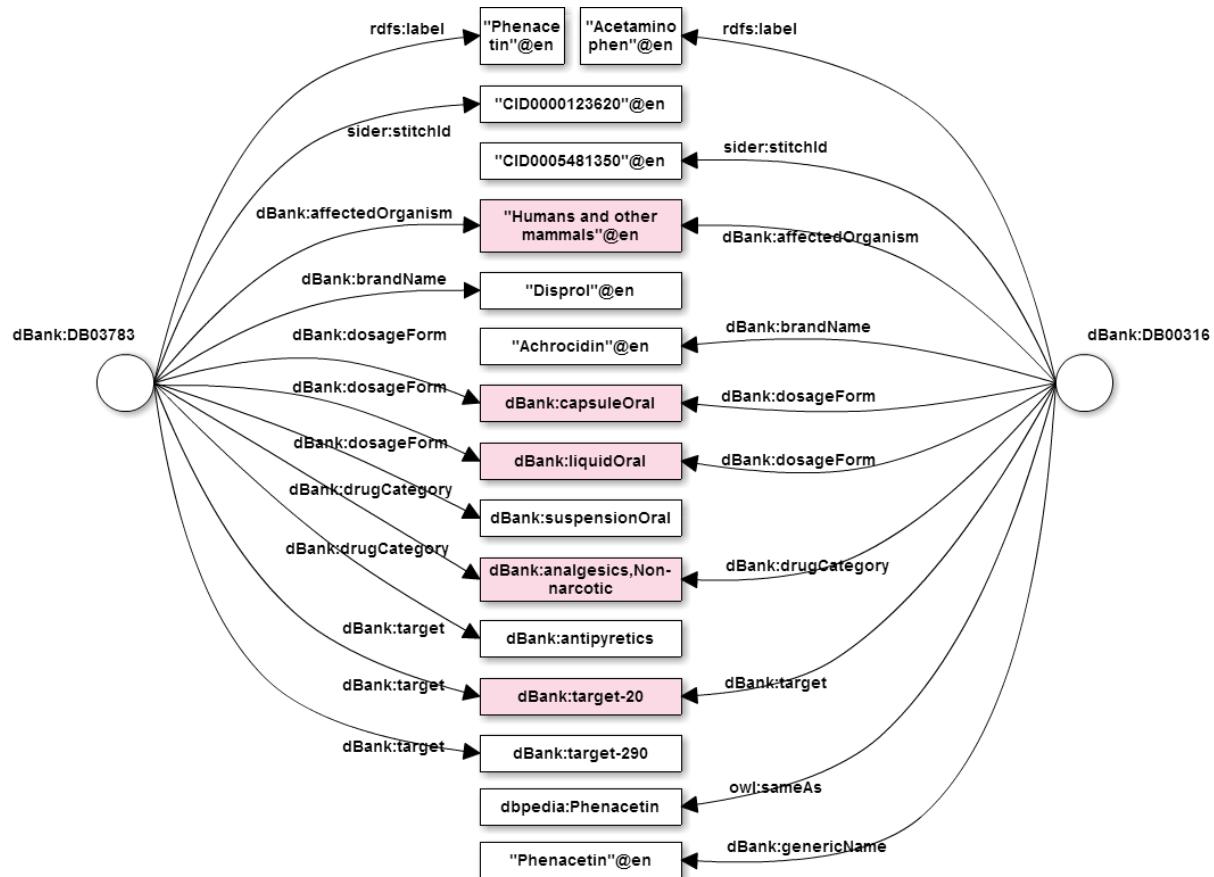


Figure 3. Subset of relations between two drugs.

Based on the relations shown in the graph, the feature vectors for `dBank:DB03783` and `dBank:DB00316` can be generated. Thus, the similarity between these two drugs is given by the 3w-Jaccard similarity of the feature vectors.

$$S_{3W\text{-JACCARD}}(\text{dBank:DB03783, dBank:DB00316}) = \frac{3 \times 5}{3 \times 5 + 6 + 5} = 0.5769$$

^eWe have experimented with other measures such as information content and Jaccard, but 3w-Jaccard performed best with the presently used data. However, combinations of different similarity measures may lead to further improvements and lower sensitivity to similarity thresholding, therefore related studies are an important aspect of our future work.

Prediction of Side Effects

All feature vectors and similarities between every pair of drugs are computed and stored in an in-memory drug similarity database (see Figure 1) optimized to support on-line querying. After this step, the off-line phase of the system is completed. Conceptually, this database represents a vector space model where each drug is represented as a dot in a n -dimensional space (where n is the maximum cardinality of the feature vectors). Given a drug x_i , the neighborhood of x_i is represented by the closest drugs (*i.e.*, the ones with higher similarity). We use k -NN algorithm¹⁷ to extract such neighborhoods. In the presented experiments we consider neighborhoods of size $k = 50$, and further filter the neighbors according to different thresholds over the similarity between $[0.0 - 1.0]$ with incremental steps of 0.1.

Side effects are propagated from one drug to its closer neighbors. That is, for a given drug x_i , the system collects the already known side effects of the drugs in the neighborhood of x_i . The side effects are combined and ranked according to their co-occurrence (frequency) to determine the link's weight. Let \mathbf{W}_{UL} be the vector with the distance of x_i to each neighbor, \mathbf{L}_{UU} be the sum of the distances from x_i to all its neighbors, $\sum_{w_i \in \mathbf{W}_{UL}} w_i$; and \mathbf{f}_L the vector of the relative frequencies for a given side effect s in all the neighbor drugs. To compute the weight that the side effect s will have over x_i , we use the average mean formula:

$$s_{weight}(x_i) = \frac{1}{\mathbf{L}_{UU}} \mathbf{W}_{UL} \mathbf{f}_L.$$

During our evaluation, for each drug in our knowledge graph, we produce as many predictions as actual side effects the drug has in SIDER, and check whether we can retrieve the same set of side effects for a given drug. Therefore, to predict the set of side effects for a given drug, we change the vector \mathbf{f}_L accordingly to each side effect in SIDER.

Results and Discussion

In this section we first describe the data and the methodology we used for evaluating our results. The results achieved are summarized then, including comparison to related works. We provide examples of selected results, and eventually discuss the benefits and drawbacks of our work in relation to the state of the art.

Evaluation Data Set and Methodology

The comparative evaluation of our approach to adverse drug reaction discovery was based on the SIDER data set as available in the Bio2RDF project. Basic statistics about the knowledge graph used in our work, is provided in Table 2.

Table 2. Basic statistics about the SIDER dataset used.

Number of drugs	731
Number of side effects (<i>i.e.</i> , ADR)	4,652
Number of drug-side effect relations	76,938
min / max / avg number of side effects per drug	1 / 771 / 105.25
min / max / avg number of drugs per side effect	1 / 631 / 16.54

We used the SIDER drug-adverse effect relationship instances as a gold standard. We performed leave-one-out cross-validation of our approach, measuring various scores to assess the predictive power of our approach by training it on a subset of the gold standard drugs and testing it on the remainder in an iterative manner. Such an evaluation method is commonly used for assessing the performance of ADR discovery systems¹⁰⁻¹⁴ and thus provides convenient means for direct comparison with related state of the art.

For the performance evaluation we use specific evaluation metrics for multi-label learning, which are different from the ones used in traditional supervised learning¹⁸. Let p be the size of the set of drugs, thus, for each drug x_i with $1 \leq i \leq p$ we have a set $Y(x_i)$ of actual side effects (from SIDER), and a set $G(x_i)$ of predicted side effects using our method. We compute the following four measures for evaluating the results regardless of the ranking of the predictions:

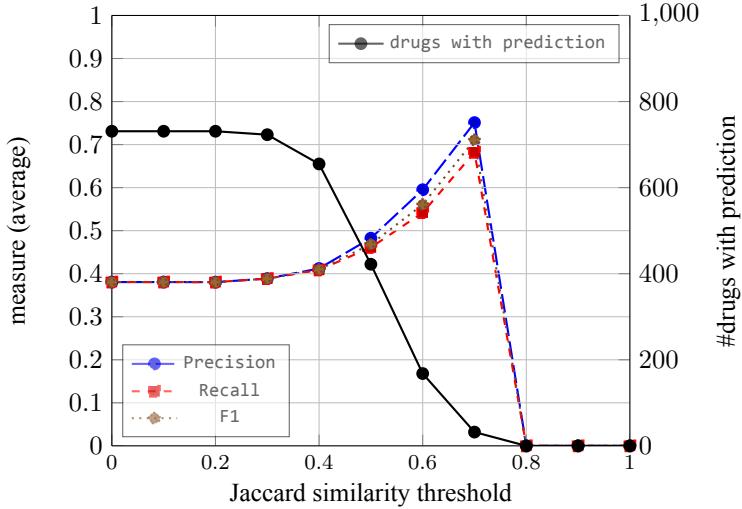


Figure 4. Plot of the results in relation to the similarity threshold.

$$\text{accuracy}(A) = \frac{1}{p} \sum_{i=1}^p \frac{|Y(x_i) \cap G(x_i)|}{|Y(x_i) \cup G(x_i)|}, \text{precision}(P) = \frac{1}{p} \sum_{i=1}^p \frac{|Y(x_i) \cap G(x_i)|}{|G(x_i)|}, \text{recall}(R) = \frac{1}{p} \sum_{i=1}^p \frac{|Y(x_i) \cap G(x_i)|}{|Y(x_i)|},$$

and F1-score, which is known as the harmonic mean of precision and recall. Furthermore, we computed three scores that reflect the prediction ranking and the extent to which the methods produce not only good, but also highly-ranked good results, namely: average precision (AP), as defined in¹⁹; TopK score, which is the relative frequency of drugs having at least one known (*i.e.*, gold standard) side-effect which is ranked among the top K high scoring side-effects according to a prediction^f; P@K score stands for the precision at K, *i.e.*, precision computed only among top K ranking side effects per drug (we used K ∈ {3, 5, 10}). We compute the average values of the particular measures across all drugs with non-empty set of ADR predictions made, which is an approach common to the works we compare ourselves to in the next section.

Among the measures we used, the TopK and P@K are arguably the most accurate scores in terms of evaluating the benefit of ADR discovery for certain types of end users like clinical practitioners. As explained in²⁰, these scores are easily grasped by non-informaticians and are therefore apt for explaining the reliability of the system to them. Moreover, in settings where quick decisions are needed, like in clinical practice, users do not tend to perform comprehensive search among many possible alternatives to find the relevant ones²⁰. The TopK and P@K scores reflect the likelihood that such users will find relevant results very quickly at the top of the list of possibly relevant results. For users who require a comprehensive study of possible ADRs (*e.g.*, pharmaceutical researchers), good evaluation measures across the whole range of predictions are important as well, though.

Summary of the Results

Figure 4 shows how the core evaluation metrics depend on the choice of the similarity cut-off threshold.

Although using a non-zero cut-off results in reduced number of drugs for which we are able to predict side effects by propagation, there are significant gains in all measures up until the 0.6 threshold. After that, the slight gain is balanced by a drastic reduction in the number of drugs with predictions, therefore we decided use 0.6 as the preferred cut-off threshold.

Table 3 compares our results achieved for the best similarity threshold (*i.e.*, 0.6) to recent related approaches (*i.e.*, those that used SIDER in a cross-validation experiment). The random baseline assigns |Y(x_i)| ADRs to all drugs x_i in a random manner, where Y(x_i) is the set of actual ADRs for the drug x_i in SIDER. The second line in Table 3 corresponds to our approach. The methods that support ranking of the side effects predicted for particular drugs are prefixed by ⁺ in the table.

^f Similarly to¹², we used K ∈ {1, 5}, but we report a relative instead of absolute frequency as it provides for a better comparison.

The method¹¹ only provides box plots with comparable measures and the value given here is the best mean as observed in the reported plots. The last row of Table 3 presents relative improvement of the presented method achieved over the best previously existing method.

Table 3. Comparison of the results with related methods.

Method	P	R	F1	AP	Top1	Top5	A
+random baseline	0.0198	0.0195	0.0196	0.057	0.0266	0.103	0.01
+Fujitsu/Insight method	0.5951	0.5419	0.5606	0.6349	0.5702	0.9532	0.4141
+Atias and Sharan (2011) ¹⁰	N/A	N/A	N/A	N/A	0.3468	0.6344	N/A
+Pauwels et al. (2011) ¹¹	N/A	N/A	N/A	N/A	N/A	N/A	ca. 0.3
+Yamanishi et al. (2012) ¹²	N/A	N/A	N/A	N/A	0.4255	0.7006	N/A
Zhang et al. (2015) ¹³	N/A	N/A	N/A	0.5134	N/A	N/A	N/A
Zhou et al. (2015) ¹⁴	0.565	0.24	0.337	N/A	N/A	N/A	N/A
Relative improvement	+5.33%	+125.79%	+66.35%	+23.67%	+34.01%	+36.05%	+38.03%

The P@K results are given in a separate Table 4 since no related approach reports these measures. The best results in both tables are highlighted in bold font.

Table 4. P@K results.

Method	P@3	P@5	P@10
random baseline	0.0179	0.0213	0.0219
Fujitsu/Insight method	0.6105	0.6239	0.6305

Examples of Results

Examples of top five results according to the F1 and P@5 measures, respectively, are given in Table 5. There is substantial overlap between the top positions of these lists, therefore we give the top five drugs for the F1 scores, and then the top five drugs out of top 30 of those that performed best at the P@5 score but were not in the top-F1 list. We can see some relatively frequent drug types as best performers, such as barbiturates, anti-histamines or NSAIDs in Table 5. This may indicate that our method can provide very good results for certain drug classes, possibly based on some of their inherent features in the Bio2RDF data. This insight is further supported by analyzing the numbers of drugs that perform well in terms of precision at 3 or 5, as depicted in Figure 5.

Table 5. Examples of top-scoring drugs.

TOP-F1				TOP-P@5			
Drug	Drug type	F1	P@5	Drug	Drug type	F1	P@5
Secobarbital	barbiturate	0.9825	1.0	Etodolac	NSAID	0.7059	1.0
Carbinoxamine	antihistamine	0.9767	1.0	Ganciclovir	antiviral drug	0.6916	1.0
Diphenhydramine	H1 histamine antagonist	0.9762	0.71	Sulindac	NSAID	0.6489	1.0
Hydroflumethiazide	diuretic	0.9697	1.0	Ketorolac	NSAID	0.6264	1.0
Pentobarbital	barbiturate	0.9643	1.0	Lansoprazole	proton pump	0.6016	1.0

The figure shows that although there is relatively large number of drugs that perform rather poorly, there is also disproportional number of drugs that perform very well.

This preliminary analysis may indicate that for some types of drugs, the results of our method can be substantially better than the average values reported above. Also, there may be strong correlations between good results among the top K results and among the whole prediction set. After more precisely detecting the types of drugs that exhibit such behavior, our method can be useful in both of the characteristic use cases we have outlined before (*i.e.*, the clinical practice vs. pharmaceutical research).

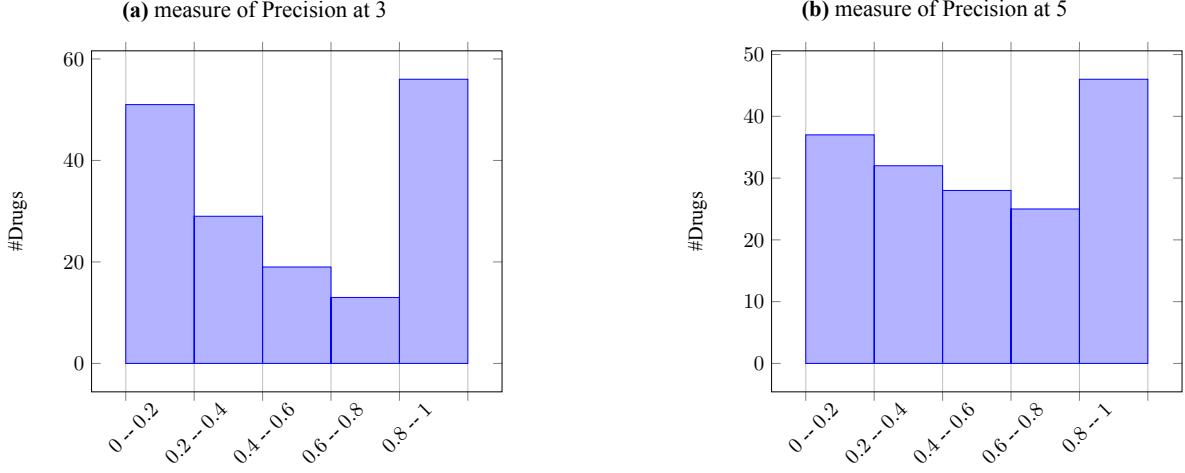


Figure 5. Histograms of drug precisions@K.

Discussion

Our results clearly show the superiority of the presented method. We achieve the best performance in all metrics, improve the previous results by up to 125.79%. The main contribution of our method over related state of the art is the combination of the following factors: (1) The presented method achieves better results in the Top1 and Top5 measures than the related approaches (by as much as 36.05%). It also provides the corresponding precision at K scores that help to determine how many of the top results are actually typically relevant (*e.g.*, more than 3 out of top 5). These combined measures allow for better assessment of what can be expected from the system in cases when users may want to process only first few results and still get many relevant predictions. (2) The method does not depend on patient or report data, even though they can be easily integrated (*e.g.*, via the ClinicalTrial.gov data set in Bio2RDF). (3) The method can automatically digest up to 12 data sets related to drugs, their structure, mechanisms of action, etc. (based on Bio2RDF status as of April 2016). This is different from related state of the art methods that would need to write specific pre-processing pipelines for each of the original data sets, or emulate our approach that is capable of using the Linked Open Data versions of biomedical databases in Bio2RDF. (4) There is only one parameter that has substantial influence on the results -- the similarity threshold. And even using the default parameter value determined by our empirical study is robust, as observed in cross-validation experiments.

The most serious limitation of our approach is the comparatively lower number of drugs for which predictions can be made with the presented method due to the drug similarity threshold. At this stage, the limitation can be partly mitigated by relaxing the similarity threshold. For instance, values around 0.2 result in predictions for all drugs we have features for, while the evaluation scores are still relatively competitive (*e.g.*, F1, Top5 and P@5 scores are 0.3804, 0.8413 and 0.4052, respectively). Focusing on the drug classes for which the presented method performs extremely well (such as NSAIDs) can lead to complete mitigation of this limitation in practical applications. For future versions of the presented technology, we have been addressing this issue by more sophisticated propagation methods involving broader neighborhoods in the similarity graph and combinations of multiple complementary similarity measures. This is expected to lead to high quality predictions for all drugs processed by the system.

Conclusions and Future Work

We presented a method for discovery of drug side effects (or, alternatively, adverse drug reactions) based on propagation of known side effects between similar drugs. The drug similarities were computed using features selected from two data sets---DrugBank and SIDER---represented in a common machine-readable format as parts of the Bio2RDF project. The presented method is very flexible in terms of adding new sets of features from other relevant resources represented in Bio2RDF, which is an advantage over many related state of the art approaches. In addition to that, we were able to achieve better results than related systems in seven standard evaluation metrics. Our system produced promising results especially relevant to use cases in which one needs high ratio of relevant results present among top few side effect predictions (*e.g.*, clinical practice where physicians may need to check for possibly dangerous side effects of a drug but do not have time to filter through many possibly relevant options). The results also motivate some interesting areas of future work that could increase practical relevance of the presented method.

The most immediate future work stemming from the research presented here addresses the sensitivity of the results to the drug similarity threshold. We aim to investigate the influence of adding additional features (from other data sets available via Bio2RDF) for the similarity computation, which may result in more densely overlapping feature space and thus higher numbers of drugs that make it even over aggressive thresholds. As more features may lead to more noise, we plan to experiment also with various feature extraction techniques to minimize the risk. We also want to experiment with multiple combined similarity measures and incorporate unrestricted propagation of the side effect labels within the computed similarity networks, using the technique first presented in²¹. This will effectively mean that a side effect may be propagated not only to direct neighbors in the drug similarity network, but also to nodes further away (if the similarity links are strong enough). Last but not least, we want to perform an in-depth analysis of the performance of our method across specific types of drugs and test its predictive capabilities on drugs not represented in SIDER, following the approach presented in¹².

Acknowledgements

This work has been supported by the TOMOE project funded by Fujitsu Laboratories Limited and Insight Centre for Data Analytics at National University of Ireland Galway. The technology described in this paper is a subject of a European patent pending, No. 15198304.6-1951.

Supplementary Material

We host all experimental data used in this paper at <http://bit.ly/AMIA2016KEDI>. The materials contain pre-processed versions of DrugBank and SIDER data sets from Bio2RDF v4.0, including fixing of a few syntax errors and a simple documentation.

References

1. Gurwitz JH, Field TS, Harrold LR, Rothschild J, Debellis K, Seger AC, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *Jama*. 2003;289(9):1107--1116.
2. Schneeweiss S, Hasford J, Göttler M, Hoffmann A, Riethling AK, Avorn J. Admissions caused by adverse drug events to internal medicine and emergency departments in hospitals: a longitudinal population-based study. *European journal of clinical pharmacology*. 2002;58(4):285--291.
3. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *The Lancet*. 2000;356(9237):1255 -- 1259.
4. Karimi S, Wang C, Metke-Jimenez A, Gaire R, Paris C. Text and Data Mining Techniques in Adverse Drug Reaction Detection. *ACM Comput Surv*. 2015 May;47(4):56:1--56:39.
5. Harpaz R, Vilar S, DuMouchel W, Salmasian H, Haerian K, Shah NH, et al. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *Journal of the American Medical Informatics Association*. 2013;20(3):413--419.
6. Wishart DS, Knox C, Guo A, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*. 2006;34(Database-Issue):668--672.
7. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Research*. 2015;
8. Belleau F, Nolin MA, Tourigny N, Rigault P, Morissette J. Bio2RDF: Towards a Mashup to Build Bioinformatics Knowledge Systems. *J of Biomedical Informatics*. 2008 Oct;41(5):706--716.
9. Nickel M, Murphy K, Tresp V, Gabrilovich E. A review of relational machine learning for knowledge graphs: From multi-relational link prediction to automated knowledge graph construction. *arXiv preprint arXiv:150300759*. 2015;.
10. Atias N, Sharan R. An algorithmic framework for predicting side effects of drugs. *Journal of Computational Biology*. 2011;18(3):207--218.
11. Pauwels E, Stoven V, Yamanishi Y. Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC bioinformatics*. 2011;12(1):169.
12. Yamanishi Y, Pauwels E, Kotera M. Drug side-effect prediction based on the integration of chemical and biological spaces. *Journal of chemical information and modeling*. 2012;52(12):3284--3292.
13. Zhang W, Liu F, Luo L, Zhang J. Predicting drug side effects by multi-label learning and ensemble learning. *BMC bioinformatics*. 2015;16(1):1.
14. Zhou H, Gao M, Skolnick J. Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Scientific reports*. 2015;5.
15. Schreiber G, Raimond Y, editors. RDF 1.1 Primer [Internet]. World Wide Web Consortium; 2014. [updated 2014 February 25, cited 2016 January 4th]. Available from: <http://www.w3.org/TR/rdf11-primer/>.

16. Choi Ss, Cha Sh, Tappert CC. A Survey of Binary Similarity and Distance Measures. *Journal on Systemics, Cybernetics and Informatics*. 2010;0(1):43--48.
17. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE transactions on information theory*. 1967;13(1):21--27.
18. Zhang ML, Zhou ZH. A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*. 2014 Aug;26(8):1819--1837.
19. Liu TY. Learning to rank for information retrieval. Springer Science & Business Media; 2011.
20. Manning CD, Raghavan P, Schütze H. 8. In: Chapter 8: Evaluation in information retrieval. Cambridge university press Cambridge; 2008. p. 151--175.
21. Zhu X, Ghahramani Z, Lafferty J, et al. Semi-supervised learning using gaussian fields and harmonic functions. In: ICML. vol. 3; 2003. p. 912--919.