Predicting Drug-Drug Interactions from Molecular Structure Images

Devendra Singh Dhami  
University of Texas, Dallas  
devendra.dhami@utdallas.edu

Gautam Kunapuli  
University of Texas, Dallas  
gautam.kunapuli@utdallas.edu

David Page  
Duke University  
david.page@duke.edu

Sriraam Natarajan  
University of Texas, Dallas  
sriraam.natarajan@utdallas.edu

Abstract

Predicting and discovering drug-drug interactions (DDIs) is an important problem and has been studied extensively both from medical and machine learning point of view. Almost all of the machine learning approaches have focused on text data or textual representation of the structural data of drugs. We present the first work that uses drug structure images as the input and utilizes a Siamese convolutional network architecture to predict DDIs.

Introduction

Adverse drug events (ADEs) are “injuries resulting from medical intervention related to a drug” (Nebeker, Barach, and Samore 2004), and are distinct from medication errors (inappropriate prescription, dispensing, usage etc.) as they are caused by drugs at normal dosages. According to the National Center for Health Statistics (NCHS 2014), 48.9% of Americans took at least one prescription drug in the last 30 days, 23.1% took at least three, and 11.9% took at least five. These numbers rise sharply to 90.6%, 66.8% and 40.7% respectively, among older adults (65 years or older). This means that the potential for ADEs is very high in a variety of health care settings including inpatient, outpatient and long-term care settings. For example, in inpatient settings, ADEs can account for as many as one-third of hospital-related complications, affect up to 2 million hospital stays annually, and prolong hospital stays by 2–5 days (DHHS 2010).

The economic impact of these issues is as widespread as the various healthcare settings and can be staggering. Estimates suggest that ADEs contributed to $3.6 billion in excess healthcare costs in the US alone (Aspden et al. 2007). Unsurprisingly, older adults are at the highest risk of being affected by an ADE, and are seven times more likely than younger persons to require hospital admission (Budnitz et al. 2006). In the US, as a large number of older adults are Medicare beneficiaries, this economic impact is borne by an already overburdened Medicare system and ultimately passed on to taxpayers and society at large. Beyond older adults, there are several other patient populations that are also vulnerable to ADEs including children, those with lower socio-economic means, those with limited access to healthcare services, and certain minorities.

Recent research has identified, somewhat surprisingly, that many of these ADEs can be attributed to very common medications (Budnitz et al. 2011) and many of them are preventable (Gurwitz and others 2003) or ameliorable (Forster et al. 2005). This issue motivates our long-term goal of developing accessible and robust means of identifying ADEs in a disease/drug-agnostic manner and across a variety of healthcare settings. Here, we focus on the problem of drug-drug interactions (DDIs), which are a type of ADE. An ADE is characterized as a DDI when multiple medications are co-administered and cause an adverse effect on the patient. DDIs, often caused by inadequate understanding of various drug-drug contraindications, are a major cause of hospital admissions, rehospitalizations, emergency room visits, and even death (Becker et al. 2007).

Identifying DDIs is an important task during drug design and testing, and regulatory agencies such as the U. S. Food and Drug Administration require large controlled clinical trials before approval. Beyond their expense and time-consuming nature, it is impossible to discover all possible interactions during such clinical trials. This necessitates the need for computational methods for DDI prediction. A substantial amount of work in DDI focuses on text-mining (Liu and Chen 2013; Chee, Berlin, and Schatz 2011) to extract DDIs from large text corpora; however, this type of information extraction does not discover new interactions, and only serves to extract in vivo or in vitro discoveries from publications.

Our goal is to discover DDIs in large drug databases by exploiting various properties of the drugs and identifying patterns in drug interaction behaviors. Recent approaches consider phenotypic, therapeutic, structural, genomic and reactive properties of drugs (Cheng and Zhao 2014) or their combinations (Dhami et al. 2018) to characterize drug interactivity. We take a fresh and completely new perspective on DDI prediction through the lens of molecular images, a few examples shown in figure 1, via deep learning. Our work is novel in the following significant ways:

- we formulate DDI discovery as a link prediction problem;
- we aim to perform DDI discovery directly on molecu-
lar structure images of the drugs directly, rather than on lossy, string-based representations such as SMILES strings and molecular fingerprints; and

• we utilize deep learning, specifically Siamese networks (Chopra et al. 2005) in a contrastive manner to build a DDI discovery engine that can be integrated into a drug database seamlessly.

Related Work

The social and economic impacts of drug-drug interactions have also been well studied and understood. The effect of DDI on medication management and social care is studied in (Arnold et al. 2018) and with its economic impact shown in (Shad, Marsh, and Preskorn 2001). The impact of DDIs in the elderly patients in 6 European countries was documented in (Björkman et al. 2002) and in a similar vein the study by Becker et al. (Becker et al. 2007) identifies that the elderly have an increased risk factor ≈ 9 times over the general population with the clinical significance of DDIs studied in (Roberts et al. 1996). Identification of DDIs can be done by either clinical trials or in vitro and in vivo experiments but these approaches are highly labor-intensive, costly and time-consuming. Thus, a system that can mitigate these factors is highly desirable.

Drug-Drug interactions have been studied extensively both from medical and machine learning point of view. From a medical standpoint (Lau et al. 2003), (Hirano et al. 2006) and (Wang et al. 2000) showed the effect of important individual drugs and enzymes such as substrates on various drug-drug interactions. The problem of DDI discovery/prediction is a pairwise classification task and thus kernel-based methods (Shawe-Taylor and Cristianini 2004) are a natural fit since kernels are naturally suited to representing pairwise similarities. Most similarity-based methods for DDI discovery/prediction have used biomedical research literature as the underlying data source and construct NLP-based kernels from these medical documents (Segura-Bedmar, Martinez, and de Pablo-Sánchez 2011; Chowdhury and Lavelli 2013). Some work has also been done on learning kernels from different types of data such as molecular and structural properties of the drugs and then using these multiple kernels to predict DDIs (Cheng and Zhao 2014; Dhami et al. 2018).

Siamese networks have been applied in one shot image recognition (Koch, Zemel, and Salakhutdinov 2015), signature verification (Bromley et al. 1994), object tracking (Bertinetto et al. 2016) and human re-identification (Varior, Haloi, and Wang 2016).

Siamese Convolutional Network for Drug-Drug Interactions

A discriminative approach for learning a similarity metric using a Siamese architecture was introduced in (Chopra et al. 2005) which maps the input (pair of inputs) into a target space such that the distance between the mappings is minimized in the target space for similar pair of examples and maximized in case of dissimilar examples. We adapt the Siamese architecture for the task of link prediction where the link is whether two drugs interact or not. Since the Siamese architecture results in a measure of similarity between the pair of given inputs it can be thresholded in order to obtain a classification. We use contrastive loss (Hadsell, Chopra, and LeCun 2006), based on a distance metric (Euclidean distance in our case), to learn a parameterized function $F$ to obtain the mapping from the input space to the target space whose minimization can result in pushing the semantically similar examples together. An important property of the loss function is that it calculated on a pair of examples. The loss function is formulated as as follows: Let $X_1$ and $X_2$ are a pair of drug images and $Y$ is the label assigned to each of the pairs. The label $Y = 0$ if the pair of drug images do not interact and $Y = 1$ if the pair of drug images interact. Also, let $D$ be the Euclidian distance between the vector of the image pairs after being processed by the underlying Siamese network and $P$ are the parameters of the function $F$. The contrastive loss
Figure 2: An overview of our model for predicting drug-drug interactions

The contrastive loss function can then be given as

$$\mathcal{L}(P, X_1, X_2, Y) = \frac{(1 - Y)}{2} D_P^2 + \frac{Y}{2} \left( \max(0, m - D_P) \right)^2$$

where $$D_P = \|F_P(X_1) - F_P(X_2)\|^2$$ is the Euclidean distance between the obtained outputs after the input pairs are processed by the sub-networks. Also $$m$$ is a margin such that $$m \geq 0$$ that signifies that dissimilar pairs beyond this margin will not contribute to the loss.

Figure 2 shows our complete architecture. It consists of two identical sub-networks i.e., networks having same configuration with the same parameters and weights. Each sub-network takes a gray-scale image of size 500 × 500 × 1 as input (we initially have color images that we convert to gray-scale before feeding to sub-networks as input) and consists of 4 convolutional layers with number of filters as 64, 128, 128 and 256 respectively. The kernel size for each convolutional layer is (9 × 9) and the activation function is \texttt{relu}. The \texttt{relu} is a non-linear activation function is given as $$f(x) = \max(0, x)$$. Each convolutional layer is followed by a max-pooling layer with pool size of (3 × 3) and a batch normalization layer. After the convolutional layers, the sub-network has 3 fully connected layers with 256, 128 and 20 neurons respectively. Thus after an image pair is processed by the Siamese sub-networks two vectors of dimension 20 × 1 are obtained. Contrastive loss is then applied to the obtained pair of vectors to obtain a distance between the input pair which can then be thresholded to obtain a prediction.

Initially we keep the threshold at 0.5 and then use precision recall curve to identify the best threshold = 0.65. Note that the convolutions in the convolutional sub-network provide translational in-variance property but rotational in-variance is also important in our problem domain. This is because isomers (one of the chiral forms) of drugs are expected to react differently when interacting with a certain drug (Nguyen, He, and Pham-Huy 2006; Chhabra, Aseri, and Padmanabhan 2013). For example, Fenfluramine and Dexfenfluramine are isomers of each other and where Fenfluramine interacts with Acebutolol but Dexfenfluramine does not (Figure 4). Another example is that the L-isomer of methorphan, Levomethorphan, is an opioid analgesic, while the D-isomer, Dextromethorphan, is a dissociative cough suppressant. We discuss our results next.

Experiments

Figure 3 shows the results for using Siamese network for predicting drug-drug interactions using drug molecular structure images. Our data set consists of images of 373 drugs downloaded from the PubChem database and generate a data set of 19936 drug pairs that interact with each other ($Y = 1$) and 47424 drug pairs that do not interact with each other ($Y = 0$). We optimize our Siamese network using the Adam optimization algorithm (Kingma and Ba 2014) with a learning rate of $5 \times 10^{-5}$. The best learning rate was obtained using line search. We also tried the Root Mean Square Propagation (RMSprop) optimizer (Hinton 2012) that is an adaptive learning rate method but the results were not encouraging and thus we make use of the Adam optimizer. As mentioned before, we also keep an initial threshold of 0.5 to obtain the predictions after obtaining a distance between pair of drug images using the Siamese convolutional network.

1https://en.wikipedia.org/wiki/Enantiopure_drug
2https://pubchem.ncbi.nlm.nih.gov/
(a) Accuracy of the network with threshold = 0.5 and epochs = 20
(b) Results with threshold = 0.65 and epochs = 20
(c) Results with threshold = 0.65 and epochs = 50

Figure 3: Results using Siamese network for predicting drug-drug interactions.

Figure 4: An example of how two isomers interact differently with a single drug.

We divide the data set into 44457 training and 22903 testing examples. To introduce rotational in-variance we rotate the drug images by 90° and 180° and use them to train and test the Siamese network. After including the rotated images the data set size increases to 88914 training and 45806 testing examples. Figure 3(a) shows the accuracy after thresholding the obtained distance at 0.5 and training the network for 20 epochs. The results show that at 0.5 threshold, if the data set includes rotated drug images the network shows better results than if the data set does not include the rotated drug images. We also tried Earth movers or Wassertein distance \(^3\) as the distance metric (Yu and Herman 2005) for the Siamese architecture but it did not make much difference in the final results.

We then pick a best threshold using a precision-recall curve such that the best intuitive trade-off between Precision and Recall can be represented and obtain a value of 0.65 as the best threshold. Figures 3(b) and 3(c) present different metrics after training the network for 20 and 50 epochs respectively. The results show that even when the number of epochs are increased the results do not vary much. We present the results with the data set including the images rotated by 180° since it gives the best accuracy result as seen in figure 3(a). Another important thing to note here is that in our problem formulation recall is the most important factor that should be considered. The simple reason is that we do not want to miss any interaction i.e. a false negative results in much more serious consequences (fatalities in patients) than false positives (monetary losses such as new clinical trials) (Dhami et al. 2018). Our network achieves a recall of \( \approx 85\% \) thereby showing the effectiveness of using a Siamese architecture for predicting DDIs. Along with the high recall we also obtain a high precision of \( \approx 75\% \) thereby showing that the Siamese architecture can also extract relevant DDIs.

An important factor also to consider here is the effect of the threshold on the obtained results. With the threshold of 0.5 the effect of rotational in-variance becomes evident as the network when trained with rotated images performs better than when trained without the rotated drug images. After we find the best threshold the effect of rotational in-variance becomes negligible.

**Conclusion and Future Work**

In this work we focus on using the molecular images of the drugs in a pairwise fashion and feeding them to a rotation-invariant Siamese architecture to predict whether two drugs interact with each other. Our evaluations on the drug images obtained from PubChem database establish the superiority of our proposed approach, which is distinct from current approaches that generally use the drug molecular structure in text format such as the Simplified Molecular Input Line Entry System (SMILES) (Weininger 1988) and SMiLes ARbitratory Target Specification (SMARTS) strings (Sayle 1997).

Combining our previous work (Dhami et al. 2018) that used different similarity measures obtained from a directed graph of known chemical reactions between drugs and enzymes, transporters and inhibitors as well as the structure of

\(^3\)https://en.wikipedia.org/wiki/Wasserstein_metric
the drugs in the form of SMILES and SMARTS strings and
the current work which uses images of the drug structure is a
natural next step. Also refining the Siamese architecture and
feeding more drug images to the network are an interesting
area of future work.

References

[Arnold et al. 2018] Arnold, R. J.; Tang, J.; Schrecker, J.; and
Hild, C. 2018. Impact of definitive drug–drug interaction
testing on medication management and patient care. Drugs-
real world outcomes.

[Aspden et al. 2007] Aspden, P.; Wolcott, J.; Bootman, J. L.;
and Cronenwett, L. R. 2007. Committee on identifying and
preventing medication errors. Preventing medication errors:
quality chasm series 1269–1272.

[Becker et al. 2007] Becker, M. L.; Kallewaard, M.; Caspers,
P. W.; Visser, L. E.; Leufkens, H. G.; and Stricker, B. H.
2007. Hospitalisations and emergency department visits due
to drug–drug interactions: a literature review. Pharmacoepi-
demiology and drug safety.

[Bertinetto et al. 2016] Bertinetto, L.; Valmadre, J.; Hen-
riques, J. F.; Vedaldi, A.; and Torr, P. H. 2016. Fully-
convolutional siamese networks for object tracking. In ECCV.

[Björkman et al. 2002] Björkman, I. K.; Fastbom, J.; Schmidt,
I. K.; Bernsten, C. B.; and of the Elderly in Europe Research
(PEER) Group, P. C. 2002. Drugdrug interactions in the
elderly. Annals of Pharmacotherapy.

[Bromley et al. 1994] Bromley, J.; Guyon, I.; LeCun, Y.;
and Shah, R. 1994. Signature verification using a” siamese”
time delay neural network. In NIPS.

[Budnitz et al. 2006] Budnitz, D. S.; Pollock, D. A.; Weiden-
bach, K. N.; Mendelsohn, A. B.; Schroeder, T. J.; and Annest,
J. L. 2006. National Surveillance of Emergency Department
Visits for Outpatient Adverse Drug Events. JAMA
296(15):1858–1866.

[Budnitz et al. 2011] Budnitz, D. S.; Lovegrove, M. C.; She-
hab, N.; and Richards, C. L. 2011. Emergency hospitali-
fizations for adverse drug events in older americans. New
England Journal of Medicine 365(21):2002–2012. PMID:
22111719.

[Chee, Berlin, and Schatz 2011] Chee, B. W.; Berlin, R.; and
Schatz, B. 2011. Predicting adverse drug events from per-
sonal health messages. In AMIA Annual Symposium Proceed-
ings.

[Cheng and Zhao 2014] Cheng, F., and Zhao, Z. 2014. Ma-
chine learning-based prediction of drug–drug interactions
by integrating drug phenotypic, therapeutic, chemical, and
genomic properties. Journal of the American Medical Infor-
matics Association.

[Chhabra, Aseri, and Padmanabhan 2013] Chhabra, N.;
Aseri, M. L.; and Padmanabhan, D. 2013. A review of drug
isomerism and its significance. International journal of
applied and basic medical research.

[Chopra et al. 2005] Chopra, S.; Hadsell, R.; LeCun, Y.; et al.
2005. Learning a similarity metric discriminatively, with
application to face verification. In CVPR (1).

[Chowdhury and Lavelli 2013] Chowdhury, M. F. M., and
Lavelli, A. 2013. Fbk-irst: A multi-phase kernel based ap-
proach for drug-drug interaction detection and classification
that exploits linguistic information. In SEM.

[Dhami et al. 2018] Dhami, D. S.; Kunapuli, G.; Das, M.;
Page, D.; and Natarajan, S. 2018. Drug-drug interaction
discovery: Kernel learning from heterogeneous similarities.
Smart Health.

[DHHS 2010] DHHS. 2010. U.S. Department of Health and Human Services, Office of Inspector General (OIG). Adverse Events in Hospitals: National Inci-
dence Among Medicare Beneficiaries, Report No.: OEI-06-
09-00090. https://oig.hhs.gov/oei/reports/
OEI-06-09-00090.pdf. [Online; accessed 21-April-
2019].

[Forster et al. 2005] Forster, A. J.; Murff, H. J.; Peterson, J. F.;
Gandi, T. K.; and Bates, D. W. 2005. Adverse drug events
occurring following hospital discharge. J Gen Intern Med
20(4):317–323.

[Gurwitz and others 2003] Gurwitz, J. H., et al. 2003. Inci-
dence and preventability of adverse drug events among older
persons in the ambulatory setting. JAMA.

[Hadsell, Chopra, and LeCun 2006] Hadsell, R.; Chopra, S.;
and LeCun, Y. 2006. Dimensionality reduction by learning
an invariant mapping. In CVPR.

[Hinton 2012] Hinton, G. 2012. Lecture 6d: a separate, adap-
tive learning rate for each connection. slides of lecture neural
networks for machine learning. Technical report, Technical
report, Slides of Lecture Neural Networks for Machine
Learning.

[Hirano et al. 2006] Hirano, M.; Maeda, K.; Shitara, Y.; and
Sugiyama, Y. 2006. Drug-drug interaction between pitavas-
tatin and various drugs via oatp1b1. Drug metabolism and
disposition.

[Kingma and Ba 2014] Kingma, D. P., and Ba, J. 2014. Adam:
A method for stochastic optimization. arXiv preprint.

[Koch, Zemel, and Salakhutdinov 2015] Koch, G.; Zemel,
R.; and Salakhutdinov, R. 2015. Siamese neural networks for
one-shot image recognition. In ICML deep learning work-
shop.

[Lau et al. 2003] Lau, W. C.; Waskell, L. A.; Watkins, P. B.;
Neer, C. J.; Horowitz, K.; Hopp, A. S.; Tait, A. R.; Carville,
D. G.; Guyer, K. E.; and Bates, E. R. 2003. Atorvastatin re-
duces the ability of clopidogrel to inhibit platelet aggregation:
a new drug–drug interaction. Circulation.

[Liu and Chen 2013] Liu, X., and Chen, H. 2013. Azdrug-
ner: an information extraction system for mining patient-
reported adverse drug events in online patient forums. In
ICSH.

[NCHS 2014] NCHS. 2014. National Center for Health
Statistics, Prescription drug use in the past 30 days, by sex,
race and Hispanic origin, and age: United States, selected
years 1988–1994 through 2011–2014. https://www.
[Nebeker, Barach, and Samore 2004] Nebeker, J. R.; Barach, P.; and Samore, M. H. 2004. Clarifying Adverse Drug Events: A Clinician’s Guide to Terminology, Documentation, and Reporting. *Annals of Internal Medicine* 140(10):795–801.

[Nguyen, He, and Pham-Huy 2006] Nguyen, L. A.; He, H.; and Pham-Huy, C. 2006. Chiral drugs: an overview. *International journal of biomedical science: IJBS*.

[Roberts et al. 1996] Roberts, J. S.; Watrous, M. L.; Schulz, R. M.; Mauch, R. P.; and Nightengale, B. S. 1996. Quantifying the clinical significance of drug-drug interactions: Scaling pharmacists’ perceptions of a common interaction classification scheme. *Annals of Pharmacotherapy*.

[Sayle 1997] Sayle, R. 1997. 1st-class smarts patterns. In *EuroMUG 97*.

[Segura-Bedmar, Martinez, and de Pablo-Sánchez 2011] Segura-Bedmar, I.; Martinez, P.; and de Pablo-Sánchez, C. 2011. Using a shallow linguistic kernel for drug–drug interaction extraction. *Journal of biomedical informatics*.

[Shad, Marsh, and Preskorn 2001] Shad, M. U.; Marsh, C.; and Preskorn, S. H. 2001. The economic consequences of a drug-drug interaction. *Journal of clinical psychopharmacology*.

[Shawe-Taylor and Cristianini 2004] Shawe-Taylor, J., and Cristianini, N. 2004. *Kernel Methods for Pattern Analysis*. Cambridge Univ. Press.

[Varior, Haloi, and Wang 2016] Varior, R. R.; Haloi, M.; and Wang, G. 2016. Gated siamese convolutional neural network architecture for human re-identification. In *ECCV*.

[Wang et al. 2000] Wang, R. W.; Newton, D. J.; Liu, N.; Atkins, W. M.; and Lu, A. Y. 2000. Human cytochrome p-450 3a4: in vitro drug-drug interaction patterns are substrate-dependent. *Drug Metabolism and Disposition*.

[Weininger 1988] Weininger, D. 1988. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*.

[Yu and Herman 2005] Yu, Z., and Herman, G. 2005. On the earth mover’s distance as a histogram similarity metric for image retrieval. In *2005 IEEE International Conference on Multimedia and Expo*.