A Recommender System Approach for Predicting Drug Side Effects

Diego Galeano, Alberto Paccanaro
Department of Computer Science
Centre for Systems and Synthetic Biology
Royal Holloway, University of London
Egham, UK
diego.galeano.2014@live.rhul.ac.uk, alberto.paccanaro@rhul.ac.uk

***Abstract*—The accurate identification of drug side effects represents a major concern for public health. We propose a collaborative filtering model for large-scale prediction of drug side effects. Our approach provides side effects recommendations for drugs to safety professionals. The proposed latent factor model relies solely on the public drug-side effect relationships from safety data. Applied to 1,525 marketed drugs and 2,050 side effect terms, we achieved an AUPRC (area under the precision-recall curve) of 0.342 in a test set, with a sensitivity of 0.73 given a specificity of 0.95, providing state-of-the-art performance in side effect prediction. We analyze the performance of the method on drug-specific Anatomical Therapeutic and Chemical (ATC) category and side effect- specific medical category of disorders. Our findings suggest that latent factor models can be useful for the early and accurate detection of unknown adverse drug events.**

***Index Terms*—drug, side effects, recommendation systems, adverse drug events, latent factor model, collaborative filtering**

I. Introduction and related work

Drug side effects represent one of the leading causes of morbidity and mortality in health care [1]. The American Institute of Medicine reported that 100,000 deaths occur annually in the U.S. from medical errors, many of which are caused by unexpected drug side effects [2]. Side effects of drugs also represent a major cost for public health. For instance, in the USA only, health care surcharge is about US $136B yearly [3]. Consequences of side effects can also affect the pharma industry. In fact, major financial setbacks could occur after a marketed drug is forced to withdraw due to deadly side effects [4]. A famous case occurred in 2004 when Merck & Co had to pull out its arthritis drug Vioxx from the market because experiments from clinical trials showed an increased risk of heart attack and stroke [5].

The detection of drug side effects starts early in the pre-marketing stage, where the molecular entity (later on, commercial drug) undergoes extensive toxicity assessment *in vitro*, *in vivo* in animal models and, after approval from the corresponding agency such as the Food and Drug Administration (FDA) in the USA, in human clinical trials [6]. In spite of this effort, many side effects are not detected until the drug hits the market. For this reason, post-marketing surveillance systems, such as the FDA-Adverse Event Reporting System, are used to follow-up the drug response in the treated population.

In silico methods for side effect identification have been proposed to help identify the most likely side effect candidates for drugs [7]–[11]. One group of methods relies on the structural chemistry of the drug and on the protein target profiles. For instance, Lounkine et al. [7] predicted drug-target-side effect associations using a similarity-ensemble approach (SEA). SEA calculates whether a molecule will bind to a target based on the chemical features it shares with those of known ligands, using a statistical model to control for random similarity [12]. Yamanishi et al. [8] predicted drug-side effect associations by an integrative approach based on kernel regression models that rely on chemical structures and target proteins. LaBute et al. [9] predicted drug-side effect associations by virtually screening drug off-target associations using molecular docking. Although these methods, that rely on molecular or protein structure, offer biological interpretability, they are often limited to the set of known structures, require considerable computational power, and generate a large amount of false positives [13].

Another group of methods frames the side effect prediction problem as infering missing links in a drug-side effect bipartite network. Cami et al. [10] proposed a missing link prediction method called Predictive Pharmacosafety Networks (PPNs). PPNs exploits the structure of the binary bipartite drug-side effect network by defining several covariates that are combined in a multivariate logistic regression model. The model outputs a probability score for each missing link in the bipartite network. The advantage of PPNs is that it relies on known drug safety relationships to make accurate predictions of missing associations. In fact, *in silico* methods can further strengthen the current drug safety toolkit by serving as hypothesis generator to drug safety professionals. The method we present in this paper, belongs to this second group of methods and we predict missing links by using a collaborative filtering approach which was developed recently for recommending movies to users.

Collaborative filtering methods became popular during the 2006 Netflix Inc. Prize Competition. Netflix released data on users preferences on movies [14]. The competition ignited an unprecedented interest and advancement in the field of recommendation systems [15], [16]. In particular, latent factor models have been shown to provide good performance

over state-of-the-art methods in movie recommendation. These methods are based on the idea that there is a common set of hidden representations that can characterize the user-movies preferences. Matrix decomposition techniques are widely used to obtain these latent factors [17].

Importantly, there is a fundamental difference between the side effect prediction and the movie recommendation problems. In movie recommendation, the fundamental assumption is that any missing value is a potential rating value, i.e. an user can watch any of the listed movies in the future. However, drugs might not be able to produce *any* possible side effect [18]. For instance, blood system drugs causes few sensory and endocrine-related side effects, whereas anti-cancer drugs produce side effects in almost all human systems [20].

In this paper, we propose a latent factor model to predict unknown drug side effects. We show that our basic matrix decomposition model outperforms PPNs. We also analyze the drug- and side effect- specific category performances. We argue that the latter analysis can be useful to safety professionals interested in performing tailored predictions of drug toxicity response. The structure of the rest of the paper is as follows. First, we start with the mathematical framework of matrix decomposition. Second, we show the evaluation procedure and introduce the comparison method. Finally, we showcase the data and the results of the predictions.

II. The latent factor model

The matrix decomposition model is based on the assumption that any $m \times n$ matrix $\mathbf{R}$ of rank $k \ll min\{m, n\}$ can be expressed in the following product form of rank-k factors [21]:

$$\mathbf{R} \approx \hat{\mathbf{R}} = \mathbf{PQ} \tag{1}$$

Here, $\mathbf{P}$ is a $m \times k$ matrix, and $\mathbf{Q}$ is a $k \times n$ matrix. In the side effect prediction problem, $\mathbf{R}$ contains binary associations for m marketed drugs and n side effect terms, i.e. $r_{uj} = 1$ if the drug u is known to cause side effect j, otherwise $r_{uj} = 0$. In this model, $\mathbf{P}$ represents the drug latent space and $\mathbf{Q}$ contains the side effects coefficients. The product $\mathbf{PQ}$ is related to the probability that a given drug produces any of the given side effects, as shown in Figure 1. The decomposition in (1) can be obtained by minimizing the functional:

$$f(\mathbf{P}, \mathbf{Q}) = \frac{1}{2}\|\mathbf{R} - \mathbf{PQ}\|_F^2 + \frac{\lambda}{2}(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2) \tag{2}$$

where $\lambda(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2)$ is an $\mathcal{L}_2$ regularization term with penalty λ added to prevent over-fitting and $\|.\|_F$ is the Frobenius norm, defined as the square root of the sum of the square of its elements. The difficulty of solving (2) is due to the fact that the functional $f(\mathbf{P}, \mathbf{Q})$ is non-convex in both $\mathbf{P}$ and $\mathbf{Q}$. Thus, gradient descent-based methods are required to approximate the solution. Thus, the derivatives of (2) are:

$$\frac{\partial f(\mathbf{P}, \mathbf{Q})}{\partial \mathbf{P}} = -(\mathbf{R} - \mathbf{PQ})\mathbf{Q}^T + \lambda \mathbf{P} \tag{3}$$

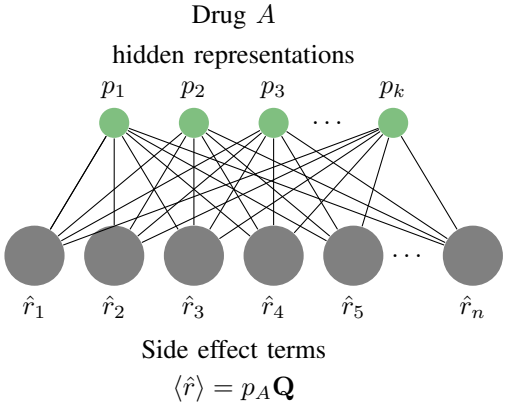

Fig. 1: The model is represented as a network depicting how the side effects in the bottom layer are generated from the Drug A hidden representations $\langle p_A \rangle = p_1, ..., p_k$ in the top layer nodes. In this illustration, p_A is a row of $\mathbf{P}$ and $\langle \hat{r} \rangle = \hat{r}_1, ..., \hat{r}_n$ is a row of $\hat{\mathbf{R}}$. $\langle \hat{r} \rangle$ contains the scores for all the side effects for the given drug.

$$\frac{\partial f(\mathbf{P}, \mathbf{Q})}{\partial \mathbf{Q}} = -\mathbf{P}^T(\mathbf{R} - \mathbf{PQ}) + \lambda \mathbf{Q} \tag{4}$$

We used conjugate gradient descent (CGD) and approximated line searches based on polynomial interpolation with Wolfe-Powel conditions to find the local minima in (2). We used a Matlab implementation of this minimizer [22]. We initialized $\mathbf{P}$ and $\mathbf{Q}$ as normally distributed random variables with small variance $\sigma^2 = 0.01$. This is done to ensure that the initial weights have small values. The learning plateaus after roughly 300 iterations (max change in the matrices between two consecutive iterations $\epsilon < 10^{-4}$).

A. The evaluation

We evaluated the performance of our method using the standard evaluation procedure for recommendation systems [21], [23]–[25]. The known entries in the $m \times n$ binary matrix $\mathbf{R}$ were randomly split into training $\mathcal{T}^{train}$ and testing $\mathcal{T}^{test}$ sets. The training set $\mathcal{T}^{train}$ contained 90% of the known entries in $\mathbf{R}$ and $\mathcal{T}^{test}$ only contained 10% of the entries. We used $\mathcal{T}^{train}$ in a ten-fold cross validation procedure for the training of the parameters λ and k. For this procedure, we randomly split the known entries in $\mathcal{T}^{train}$ into ten disjoint sets. For each set s, we trained the model with $s - 1$ sets and tested on the remaining set. The optimal parameters (k_{op}, λ_{op}) are those that maximizes the mean AUROC for the ten folds. Finally, the model performance was measured on $\mathcal{T}^{test}$ by training on the entire $\mathcal{T}^{train}$ using the optimal parameters.

B. Predictive pharmacosafety networks (PPNs)

In order to make this paper self-contained, we introduce in this subsection the state-of-the-art side effect prediction method Predictive Pharmacosafety Networks (PPNs) [10] which we use later for comparison. PPNs begins by creating a bipartite graph in which one set of nodes represent the drugs

and the other set of nodes represent the side effects. The set of edges corresponds to the known drug-side effect associations. PPNs then computes covariate measures in this network to capture the structure of connections between drugs and side effects. The covariate $X_s(u,j)$ quantify a relationship between a drug node u and a side effect node j. The predictive model is based on the binary response variable $Y_{uj}, u \in \{1,...,m\}, j \in \{1,...,n\}$, for m drugs and n side effects. Y_{uj} denotes the presence or absence of drug-side effect associations. Using a multivariate logistic-regression model, the response is modeled as a Bernoulli random variable with the following expectation

$$\mathbb{E}[Y_{uj}] = \frac{1}{1+\exp\left(-\sum_s \beta_s X_s(u,j)\right)} \tag{5}$$

where β_s denotes the model parameter and X_s the model covariate. The optimization is performed by assuming independence between the responses Y_{ij} and performing model fitting by maximum likelihood. Several covariates are considered in the PPN-NET model. The *degree covariates* are defined as follow;

$$\begin{aligned} X_1(u,j) &= deg(u) \times deg(j) \\ X_2(u,j) &= |deg(u) - deg(j)| \end{aligned} \tag{6}$$

Here, $deg(u)$ denotes the degree of node u. The degree product covariate, $X_1(u,j)$, aimed to capture preferential attachment among high-degree drugs and side effects. The degree difference covariate, $X_2(u,j)$, aimed to capture assortativity, i.e. whether high-degree drugs connects to high-degree side effects or to small-degree side effects. The *distance covariates* are also defined. These depends not only on the nodes (u,j) but on the sets of neighbor $N(u)$ and $N(j)$. Let $J(j,k)$ denotes the Jaccard similarity between the neighbors sets $N(j)$ and $N(k)$,

$$J(j,k) = \frac{|N(j) \cap N(k)|}{|N(j) \cup N(k)|} \tag{7}$$

The following Jaccard-based covariates quantify structural similarity between drug pairs and side effect pairs,

$$\begin{aligned} X_3(u,j) &= \max_{k \in N(u)-\{j\}} \{J(j,k)\} \\ X_4(u,j) &= \max_{k \in N(j)-\{u\}} \{J(u,k)\} \end{aligned} \tag{8}$$

Finally, Jaccard-based predictors based on Kullback-Leibler divergence ($\mathcal{K}_\mathcal{L}$) between the overall distribution of similarities between a drug ($\bar{D}_{se}$) and the drugs in its local neighborhood ($D_{se}(u,j)$) or between side effects ($\bar{D}_{se}$) and side effects in its neighborhood ($D_{se}(u,j)$) are defined;

$$\begin{aligned} X_5(u,j) &= \mathcal{K}_\mathcal{L}(D_{se}(u,j), \bar{D}_{se}) \\ X_6(u,j) &= \mathcal{K}_\mathcal{L}(D_{drug}(u,j), \bar{D}_{drug}) \end{aligned} \tag{9}$$

III. The Data

We used the Side effect Resource (SIDER) version 4.1. SIDER contains information on marketed medicines and their recorded side effects [26]. The information comes from public documents and package inserts collected from the Food and Drug Administration (FDA), Health Canada, European Medicines Agency (EMA), etc. The version 4.1 contains more than 1,500 marketed drugs and 5,868 side effects whose terms are mapped in the Medical Dictionary for Regulatory Activities (MedDRA). The database integrates data from pre- and post-marketing side effect evidence. A pictorial representation of the data is presented in Figure 2 which shows the twenty most popular side effects of marketed drugs.

Following Cami et al. [10], we used the drug-side effect as binary associations. First, we only considered the side effect terms that were MedDRA preferred terms (PT). Then, we filtered out the drugs that have less than 5 side effects and the side effect terms that were in less than 5 drugs. This step is important because the method requires some data for each drug and side effect to make confident prediction. Similar filtering is performed when building datasets for movie recommendations, where only users that have watched at least 20 movies are considered[1]. Thus, we obtained a gold-standard data set that contains 1,525 marketed drugs and 2,050 side effects with 148,705 known associations (i.e. density 4.75%).

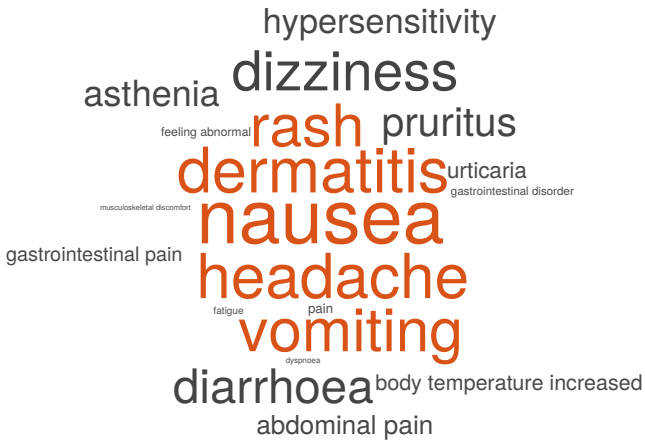

Fig. 2: WordCloud of the most popular side effect terms. The 20 most popular side effects are shown. The size of the word is proportional to the popularity of the side effect across the corpus of drugs. The five most popular side effects are indicated in orange.

IV. Results

In this section we present the results of the experiments with the proposed model. We followed the ten-fold cross

[1] https://grouplens.org/datasets/movielens/

validation procedure to find the optimal values for the regularization term (λ) and the number of hidden representations (k). We run the training with parameter values in the set $k \in \{1, 5, 10, 15, 20, 25, 30, 35, 40, 50, 100\}$ and $\lambda \in \{0.1, 1, 5, 10, 15, 20\}$. The training took roughly four hours in a computer cluster of 12 cores (Intel(R) Zeon(R) CPU E5-1650 v3 @ 3.5GHz). We found that the optimal number of parameters were ($k_{op} = 40, \lambda_{op} = 15$).

A. Latent factor models outperform PPNs

We trained our latent factor model and the network-based method PPN-NET with 90% of the data and tested on the remaining 10% of the data, i.e. 14,871 unseen drug-side effect associations using the optimal parameters. Figure 3 shows the precision-recall curve for the methods. Our latent factor model outperforms PPNs in precision-recall in Area Under the Precision-Recall Curve (AUPRC). Our model achieved 0.342 in AUPRC whereas PPN-NET achieved 0.282. In Area Under the ROC Curve (AUROC), both models achieved 0.952. Our model provides sensitivity of 0.73 at a specificity of 0.95. The curve also shows that our model performs better in regions of high precision. For instance, we achieved a recall of 20% at 70% precision, 10% higher than PPN-NET. This is particularly important for ranked recommendations of side effects. Furthermore, AUPRC is sometimes a more useful measure when there the dataset is imbalanced [8], [27].

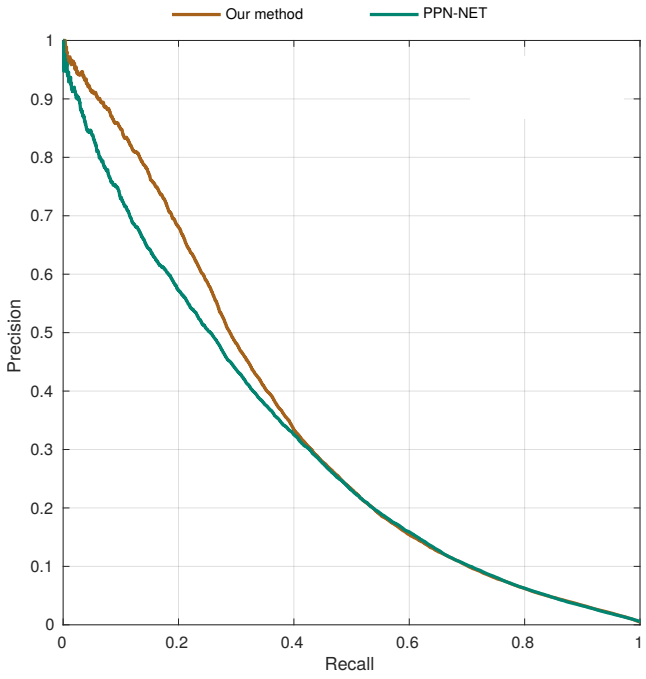

Fig. 3: Precision-recall (PR) curve of the proposed method and PPN-NET on the test set $\mathcal{T}^{test}$. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved.

To understand the performance of our method in the binary classification problem of identifying true from false drug-side effect associations, we plotted in Figure 4 the distributions of the predicted scores with respect to the binary classes for the unseen drug-side effect associations in the test set $\mathcal{T}^{test}$. The forty latent variables were able to capture complex hidden relationships between the drugs and side effects that allowed to fill in the incomplete matrix $\mathbf{R}$. These latent features can be interpreted as a low-dimensionality reduction on the original associations. Clearly, the scores for the true associations were significantly higher than for the false associations (t-test significant, $P < 4.94 \times 10^{-324}$).

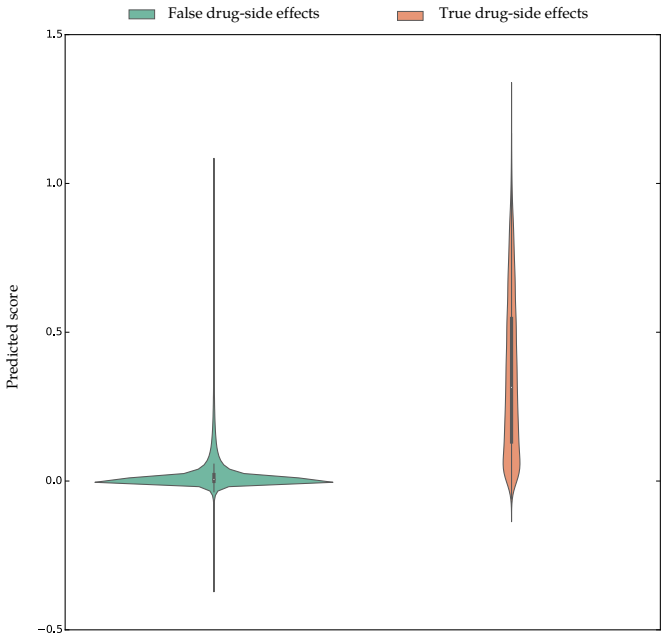

Fig. 4: Comparative violin plot of the predicted scores versus the classes in the test set $\mathcal{T}^{test}$. The classes represent the true and false drug-side effects associations. In both cases, the predicted score for the true associations is higher than for the false associations. For all side effects, mean and standard deviation 0.361 ± 0.261 (true associations) vs 0.0396 ± 0.114 (false associations).

B. Drug- and side effect-specific predictions

An interesting question is whether the prediction performance depends on drug- or side effect-specific category [10]. Drugs were classified using their Anatomical Therapeutic and Chemical (ATC) categories. These categories are controlled and curated by the World Health Organization (WHO)[2]. In the ATC system, drugs are classified in a hierarchy of five different levels. The system has fourteen main anatomical and pharmacological groups on the first level, as shown in Table I. We observed that the performance varies across therapeutic categories (Figure 5), particularly higher for systemic hormonal preparations, insulin (H) drugs and lower for Blood system (B) drugs. The drug-specific category performance implies that it is more difficult to obtain accurate ranked recommendation of side effects with high precision for certain

[2]https://www.whocc.no/atc_ddd_index/

therapeutics more than others, such as Blood system drugs (40% precision at 20% recall) and dermatological drugs (52% precision at 20% recall).

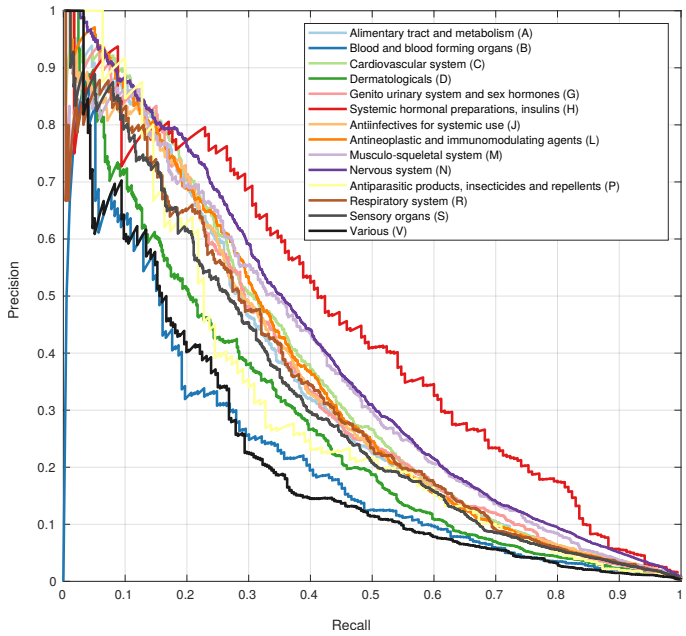

Fig. 5: Precision-recall curve for drug-specific Anatomical Therapeutic and Chemical (ATC) category. Only first level ATC categories are shown. Different drugs have different anatomical and pharmacological mode of action as represented in this level.

TABLE I: Performance by drug-specific category

Top ***ATC category***	**Performance** ***AUPRC***
Alimentary tract and metabolism (A)	0.340
Blood and blood forming organs (B)	**0.218**[a]
Cardiovascular system (C)	0.358
Dermatologicals (D)	0.274
Genito urinary system and sex hormones (G)	0.346
Systemic hormonal preparations, insulins (H)	**0.442**[b]
Antiinfectives for systemic use (J)	0.340
Antineoplastic and immunomodulating agents (L)	0.354
Musculo-squeletal system (M)	0.369
Nervous system (N)	0.397
Antiparasitic products, insecticides and repellents (P)	0.299
Respiratory system (R)	0.328
Sensory organs (S)	0.315
Various (V)	**0.218**[a]

[a,b]Lowest and highest performance.

Side effects were classified according to their top MedDRA category of disorders. MedDRA is a rich and highly specific standardised medical terminology to facilitate sharing of regulatory information internationally for medical products[3]. It is curated by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). We analyze the side effect-specific MedDRA category as shown in Table II. We observed that the prediction performance for side effect categories varies as well. The highest performance was obtained for gastrointestinal disorders whereas the lowest performance was obtained for side effects related to neoplasms benign, malignant and unspecified (including cysts and polyps).

To understand better the variability in performance by side effect category, we show in Figure 6 a representative set of MedDRA categories. As mention before, gastrointestinal side effects are easier to recall with high precision, i.e. by 70% precision at 30% recall, whereas neoplasms represent a somewhat harder problem, by 10% precision at 30% recall. Interestingly, disorders related to the immune system such as autoimmune disorders including nervous, muscular, hepatic and lupus-associated conditions achieved a relatively good performance (40% precision at 30% recall).

TABLE II: Performance by side effect-specific category

Top ***MEDDRA category***	**Performance** ***AUPRC***
Hepatobiliary disorders	0.170
Metabolism and nutrition disorders	0.275
Eye disorders	0.187
Investigations	0.192
Musculoskeletal and connective tissue disorders	0.351
Gastrointestinal disorders	**0.448**[b]
Social circumstances	0.245
Immune system disorders	0.297
Reproductive system and breast disorders	0.186
Neoplasms	**0.064**[a]
General disorders and administration site conditions	0.432
Endocrine disorders	0.163
Vascular disorders	0.332
Blood and lymphatic system disorders	0.272
Skin and subcutaneous tissue disorders	0.381
Congenital, familial and genetic disorders	0.185
Infections and infestations	0.303
Respiratory, thoracic and mediastinal disorders	0.301
Psychiatric disorders	0.352
Renal and urinary disorders	0.269
Ear and labyrinth disorders	0.358
Cardiac disorders	0.363
Nervous system disorders	0.370
Injury, poisoning and procedural complications	0.122

[a,b]Lowest and highest performance.

V. Conclusion and Discussion

In silico side effect prediction can help safety professionals to prioritize candidates in pharmacoepidemiological studies and to anticipate possible lethal outcomes [1], [10], [19]. It has shown already useful in toxicology and drug safety, but there are still many unanswered questions in the emerging field of computational pharmacology [28], such as the characterization of the incomplete drug-target networks and the prediction of lethal polypharmacy side effects [1]. The first is important to understand the soff-target proteins that cause the side effect [19], whereas the latter because most treatments consist of drug combinations that can lead to unseen and unpredictable side effects.

[3]https://www.meddra.org/

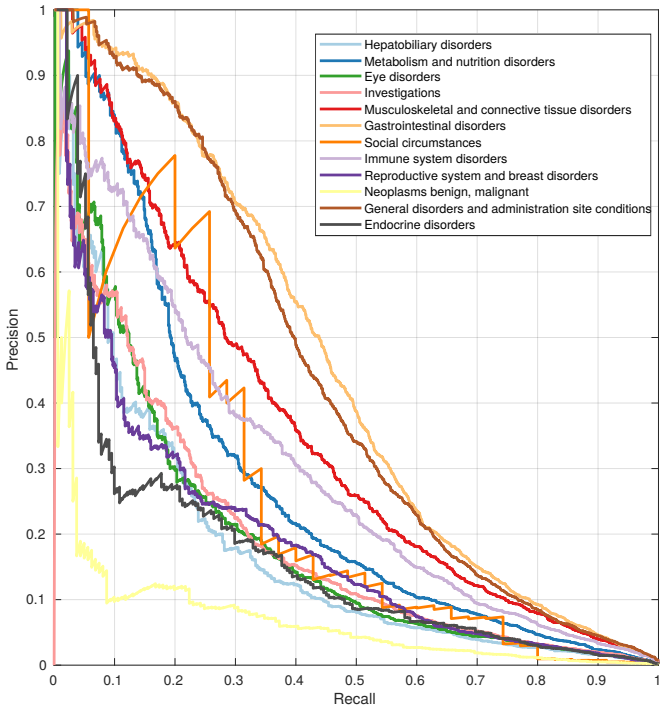

Fig. 6: Precision-recall curve for different side effect-specific MedDRA category. Only a representative number of top MedDRA disorders are shown.

In this paper, we investigated the use of collaborative filtering models for predicting side effects of marketed drugs. We trained a latent factor model on safety data from public package inserts that included 1,525 marketed drugs and 2,050 side effect unique terms. The proposed model achieved an AUPRC of 0.342 and an AUROC of 0.952 in a hold-out test set, with a sensitivity of 0.73 given a specificity of 0.95. Our findings suggest that latent factor models can be useful for predicting side effects of marketed drugs. Our method offers an alternative and efficient approach for exploring drug safety profiles on marketed drugs. We showed that our method outperforms the network-based method PPNs in AUPRC, which is of particular importance when using ranked recommendations of drug side effects. Furthermore, we analyzed the performances on drug-and side effect-specific categories. We found that there is better performance for some categories than for others.

There is still room for improvement in predicting and understanding drug side effects in the emerging field of system pharmacology. Importantly, *in silico* side effect prediction methods can be used as a complement to the wide range of *in vitro* or *in vivo* approaches to anticipate and detect toxicological liabilities in pre-and post-marketing drug development.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. A.P. and D.G. were supported in part by Biotechnology and Biological Sciences Research Council (BBSRC), grants BB/K004131/1, BB/F00964X/1 and BB/M025047/1 to A.P., CONACYT Paraguay Grant INVG01-112 (14-INV-088) and PINV15-315 (14-INV-088), and NSF Advances in Bio Informatics grant 1660648.

References

[1] N. P. Tatonetti, P. Y. Patrick, R. Daneshjou, and R. B. Altman, "Data-driven prediction of drug effects and interactions," *Science translational medicine*, vol. 4, no. 125, pp. 125ra31–125ra31, 2012.

[2] J. H. Gurwitz, T. S. Field, J. Avorn, D. McCormick, S. Jain, M. Eckler, M. Benser, A. C. Edmondson, and D. W. Bates, "Incidence and preventability of adverse drug events in nursing homes," *The American journal of medicine*, vol. 109, no. 2, pp. 87–94, 2000.

[3] F. R. Ernst and A. J. Grizzle, "Drug-related morbidity and mortality: updating the cost-of-illness model," *Journal of the American Pharmaceutical Association (1996)*, vol. 41, no. 2, pp. 192–199, 2001.

[4] H. C. J. . A. J. K. Onakpoya, I. J., "Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature." *BMC Medicine*, 2016.

[5] C. Vanchieri, "Vioxx withdrawal alarms cancer prevention researchers," *Journal of the National Cancer Institute*, vol. 96, no. 23, pp. 1734–1735, 2004.

[6] Dearden, "In silico prediction of drug toxicity," *Comput Aided Mol Des*, 2003.

[7] E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté *et al.*, "Large-scale prediction and testing of drug activity on side-effect targets," *Nature*, vol. 486, no. 7403, pp. 361–367, 2012.

[8] Y. Yamanishi, E. Pauwels, and M. Kotera, "Drug side-effect prediction based on the integration of chemical and biological spaces," *Journal of chemical information and modeling*, vol. 52, no. 12, pp. 3284–3292, 2012.

[9] M. X. LaBute, X. Zhang, J. Lenderman, B. J. Bennion, S. E. Wong, and F. C. Lightstone, "Adverse drug reaction prediction using scores produced by large-scale drug-protein target docking on high-performance computing machines," *PloS one*, vol. 9, no. 9, p. e106298, 2014.

[10] A. Cami, A. Arnold, S. Manzi, and B. Reis, "Predicting adverse drug events using pharmacological network models," *Science translational medicine*, vol. 3, no. 114, pp. 114ra127–114ra127, 2011.

[11] I. Shaked, M. A. Oberhardt, N. Atias, R. Sharan, and E. Ruppin, "Metabolic network prediction of drug side effects," *Cell systems*, vol. 2, no. 3, pp. 209–213, 2016.

[12] M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijer, R. C. Matos, T. B. Tran *et al.*, "Predicting new molecular targets for known drugs," *Nature*, vol. 462, no. 7270, p. 175, 2009.

[13] J. T. Dudley, T. Deshpande, and A. J. Butte, "Exploiting drug–disease relationships for computational drug repositioning," *Briefings in bioinformatics*, vol. 12, no. 4, pp. 303–311, 2011.

[14] R. M. Bell and Y. Koren, "Lessons from the netflix prize challenge," *Acm Sigkdd Explorations Newsletter*, vol. 9, no. 2, pp. 75–79, 2007.

[15] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, 2009.

[16] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. Ieee, 2008, pp. 263–272.

[17] P. Cremonesi, Y. Koren, and R. Turrin, "Performance of recommender algorithms on top-n recommendation tasks," in *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 2010, pp. 39–46.

[18] Y. Kwon, *Handbook of essential pharmacokinetics, pharmacodynamics and drug metabolism for industrial scientists*. Springer Science & Business Media, 2001.

[19] M. Campillos, M. Kuhn, A.-C. Gavin, L. J. Jensen, and P. Bork, "Drug target identification using side-effect similarity," *Science*, vol. 321, no. 5886, pp. 263–266, 2008.

[20] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, and P. Bork, "A side effect resource to capture phenotypic effects of drugs," *Molecular systems biology*, vol. 6, no. 1, p. 343, 2010.
[21] C. C. Aggarwal, *Recommender systems*. Springer, 2016.
[22] Rassmusen. (2006, 04) Minimize. [Online]. Available: http://learning.eng.cam.ac.uk/carl/code/minimize/
[23] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5–53, 2004.
[24] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, vol. 2009, p. 4, 2009.
[25] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*. Springer, 2011, pp. 1–35.
[26] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, "The sider database of drugs and side effects," *Nucleic acids research*, vol. 44, no. D1, pp. D1075–D1079, 2015.
[27] N. Atias and R. Sharan, "An algorithmic framework for predicting side effects of drugs," *Journal of Computational Biology*, vol. 18, no. 3, pp. 207–218, 2011.
[28] A. L. Hopkins, "Network pharmacology," *Nature biotechnology*, vol. 25, no. 10, p. 1110, 2007.