Boosting Collaborative Filters for Drug-Target Interaction Prediction

Cristian Orellana M.[1(✉)], Ricardo Ñanculef[1], and Carlos Valle[2]

[1] Department of Informatics,
Federico Santa María Technical University, Valparaíso, Chile
cristian.orellanam@alumnos.usm.cl, jnancu@inf.utfsm.cl
[2] Department of Computer Science and Informatics,
University of Playa Ancha, Valparaíso, Chile
carlos.valle@upla.cl

Abstract. In-silico prediction of interactions between drugs and proteins has become a crucial step in pharmaceutical sciences to reduce the time and cost required for drug discovery and repositioning. Even if the problem may be approached using standard recommendation algorithms, the accurate prediction of unknown drug-target interactions has shown to be very challenging due to the relatively small number of drugs with information of their target proteins and viceversa. This issue has been recently circumvent using regularization methods that actively exploit prior knowledge regarding drug similarities and target similarities. In this paper, we show that an additional improvement in terms of accuracy can be obtained using an ensemble approach which learns to combine multiple regularized filters for prediction. Our experiments on eight drug-protein interaction datasets show that most of the time this method outperforms a single predictor and other recommender systems based on multiple filters but not specialized to the drug-target interaction prediction task.

Keywords: Drug-target interaction prediction · Collaborative filtering · Ensemble methods

1 Introduction

Discovering novel drug-target interactions (DTI) is one of the fundamental tasks in pharmaceutical sciences [3]. As in-vivo experimental methods are extremely costly and time-consuming, computational approaches capable to select the most promising candidates for a further validation have become of great importance in the last years [3,9]. From a machine learning perspective, a DTI problem can be approached as a recommendation task, where for a given drug (or target) a ranking of "expected" target proteins (or drugs) interactions is generated.

Similar to implicit recommendation tasks, in DTI problems, only sparse information for interacting pairs is available. That means one cannot assume

R. Vera-Rodriguez et al. (Eds.): CIARP 2018, LNCS 11401, pp. 212–220, 2019.
https://doi.org/10.1007/978-3-030-13469-3_25

non-interacting pairs as truly negatives, because some of them correspond to interacting pairs not discovered yet. DTI prediction is also challenging because publicly available databases contain a extremely small amount of validated pairs [8]. Even if sparsity is also a challenging feature of other recommendation tasks, most settings assume a minimum of $s > 1$ annotations exist for each user and item. This context suggests that the use of knowledge beyond the known interactions may be of crucial importance to successfully address DTI problems.

Contribution. In this paper, we investigate the use of AdaBoost for DTI. Relying on a probabilistic formulation for DTI, it is possible to obtain a principled ensemble algorithm that learns to combine predictions to produce more accurate recommendations. This idea is in line with previous contributions in the collaborative filtering and DTI literature [12,13]. However, up to our knowledge, we are the first to study the use of Adaboost to build an ensemble of collaborative filters for DTI. Previous methods are based on other ensemble paradigms (e.g. stacking [15]) without collaborative filters or do not employ DTI methods as base learners (e.g. decision trees [10]). In a nutshell, our method consists in solving re-weighted versions on an objective function that has been successfully used by NRLMF, a state-of-the-art method for DTI. Our experiments on standard benchmarks show that, in general, the proposed method outperforms a single predictor and an ensemble method not specialized for DTI.

The rest of this article is organized as follows. In Sect. 2 we formalize the DTI problem and briefly discuss related work. In Sect. 3, we formulate our ensemble method. In Sect. 4 we present experimental results that demonstrate the performance of our algorithm on eight DTI datasets. Section 5 closes the article with the conclusions and final remarks.

2 Problem Statement and Related Work

Problem Definition. Given a set of drugs $D = \{d_i\}_{i=1}^{m} \subset \mathcal{D}$, a set of target proteins $T = \{t_j\}_{j=1}^{n} \subset \mathcal{T}$ and a binary matrix $R^* \in \mathbb{R}^{m \times n}$ where $R^*_{ij} = 1$ if and only if drug d_i interact with target t_j, a DTI problem consists in predicting R^* from a matrix $R \in \mathbb{R}^{m \times n}$ where some interactions has been removed, that is, $R_{ij} = 0$ but $R^*_{ij} = 1$. This definition implies that the negative examples in R are only implicit, in the sense that $R_{ij} = 0$ can represent either an interacting pair not yet discovered or a truly non-interacting pair. Besides the interaction matrix R, similarity information regarding drugs and targets can be available. This information is encoded into the form of similarity matrices $S^{(d)} \in \mathbb{R}^{m \times m}$ and $S^{(t)} \in \mathbb{R}^{n \times n}$ where a high value $S^{(d)}_{kl}$ (respectively $S^{(t)}_{kl}$) represents a high similarity between drugs d_k and d_l (respectively proteins t_k and t_l).

Related Work. Compared to more traditional recommendation tasks, DTI problems are challenging because publicly available databases contains a very small amount of validated pairs [8]. This context explains why many state of the art methods rely on knowledge beyond the partially observed matrix R. Indeed, many machine learning approaches integrating information on drug or target

similarity have been investigated in the last years. In [5], Gönen proposed a Bayesian matrix factorization model which encodes chemical similarity between compounds and genomic similarity between proteins using kernels functions [11]. This allow to perform predictions for drugs/proteins without annotations (cold start). Zheng proposed in [16] a matrix factorization approach able to accept more than one similarity matrix over drugs, as well as over targets, and able to automatically learn weights over the multiple similarity matrices in order to fit the latent matrix factors. Cobanoglu et al. adapted in [2] a probabilistic matrix factorization to DTI, demonstrating that this technique allowed to identify functionally similar drugs even in the absence of 3D shape similarity. More recently, Liu et al. has extended in [9] logistic matrix factorization (LMF) [6] to more actively exploit drug/target similarities. Neighborhood-based regularizers are incorporated into the objective function in order to constraint the latent factors of similar drugs/targets to be similar. This method, referred to as NRLMF, is shown to outperform state-of-the-art methods, including [5] and [16].

Aside matrix factorization techniques, some ensemble methods have been used for DTI problems. One of them is DrugE-Rank [15], which trains multiple similarity-based methods and use each output as feature to train a ranking learner for DTI predictions (stacking ensemble). Another ensemble approach for DTI is formulated in [10], where a boosting framework is utilized to combine multiple features for drug-target pairs, using decision trees as base learners.

A Probabilistic Model for DTI. LMF and NRMLF rely on a probabilistic model for DTI. LMF decomposes the interaction matrix R as the product $R = UV^T$ of two latent matrices $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{n\times r}$. Each row of U, u_i, encodes a latent representation for drug d_i, whereas each row of V, v_j encodes a latent representation for protein t_j. While standard matrix factorization methods models the interaction between a drug d_i and target t_j using a score $s_{ij} = u_i v_j^T$, LMF models the probability of interaction p_{ij} for a pair (d_i, t_j) using the model $p_{ij} = \sigma(s_{ij})$, where $\sigma(\xi) = \exp(\xi)/(1 + \exp(\xi))$ is the sigmoid function. The latent matrices U, V are learnt from data by maximizing the log-likelihood

$$\ell_0(U,V) = \sum_{i,j} cR_{ij} \log p_{ij} + (1 - R_{ij}) \log(1 - p_{ij}) - \frac{\lambda_d}{2}\|U\|_F^2 - \frac{\lambda_t}{2}\|V\|_F^2\,, \quad (1)$$

where $c \in \mathbb{R}$ is a parameter controlling the relative importance of positive versus negative examples and $\lambda \in \mathbb{R}$ is a regularization parameter enforcing sparsity in the latent representations. Liu et al. [9] propose to regularize this objective function in such a way that similar proteins/drugs obtain similar latent representations. Let $N(d_i)$ be the set of k_1 nearest neighbours of drug d_i computed according to the similarity matrix $S^{(d)}$ and $N(t_j)$ be the set of k_2 nearest neighbours of protein t_j according to the similarity matrix $S^{(t)}$. The new objective takes the form

$$\ell_1(U,V) = \ell_0(U,V) - \alpha/2 \sum\nolimits_{i,j} a_{ij}\|u_i - u_j\|_2^2 - \beta/2 \sum\nolimits_{i,j} b_{ij}\|v_i - v_j\|_2^2\,, \quad (2)$$

where $a_{ij} = s_{ij}^{(d)}$ if $d_j \in N(d_i)$, $a_{ij} = 0$ if $d_j \notin N(d_i)$, $b_{ij} = s_{ij}^{(t)}$ if $t_j \in N(t_i)$, $b_{ij} = 0$ if $t_j \notin N(t_i)$ and α, β are new regularization parameters. The obtained objective function is differentiable and can be optimized using gradient ascent.

3 Proposed Method

In this section we formulate an ensemble method specialized for DTI. Essentially, we demonstrate that a principled way to combine DTI filters consists in solving weighted versions of the objective function used by LMF or NRMLF.

Boosting Procedure. Let $P : \mathcal{D} \times \mathcal{T} \to [0, 1]$ be the probability distribution generating pairs (d, t) from $\mathcal{D} \times \mathcal{T}$ and $Y(d, t)$ a binary random variable such that $Y(d, t) = 1$ if (d, t) is an interacting pair and $Y(d, t) = -1$ otherwise. We cast the problem of learning an ensemble $F(d, t)$ of DTI filters as that of approximating the logit of the interaction probability, i.e.,

$$F(d,t) = \log \frac{P(Y=1|d,t)}{1-P(Y=1|d,t)} = \log P(Y=1|d,t) - \log P(Y=-1|d,t). \quad (3)$$

It is well known in machine learning that a way to obtain a hypothesis of the previous form consists in training a learner to minimize the following objective

$$J(F) = \mathbb{E}\left\{Q(Y(d,t), F(d,t))|d,t\right\}, \quad (4)$$

where $Q(Y(d, t), F(d, t)) = \exp(-Y(d, t)F(d, t))$ is known as the exponential loss. To optimize (4) we can adopt the stage-wise approach characteristic of boosting algorithms, i.e. we can implement F using an additive model of the form $F^{(k)} = \sum_{\ell}^{k} f^{(\ell)}$, where each $f^{(\ell)} : \mathcal{D} \times \mathcal{T} \to [-1, 1]$ is a DTI filter, and train $f^{(1)}, f^{(2)}, \ldots$ one after the other to improve the value of the objective function $J(F)$. It is not difficult to show indeed (see e.g. [4]) that taking a gradient descent step to expand $F^{(k)} = \sum_{\ell=1}^{k} f^{(\ell)}$ at a given iteration k correspond to choose

$$f_*^{(k+1)}(d,t) = \frac{1}{2} \log \frac{P^{(k)}(Y=1|d,t)}{(1-P^{(k)}(Y=1|d,t))}, \quad (5)$$

where $P^{(k)}(Y = 1) = W^{(k)}(d, t)P(Y(d, t) = 1)$ and

$$W^{(k)}(d,t) \propto \exp(-Y(d,t)F^{(k)}(d,t)), \quad (6)$$

represents a weighting distribution enforcing the hypothesis $f^{(k+1)}$ built at step k of the boosting procedure to focus on drug-target pairs (d, t) that the ensemble $F^{(k)}$ has incorrectly identified or has identified with a small "margin" $\eta(d, t) = Y(d, t)F^{(k)}(d, t)$. In order to implement the hypothesis in (5), we can first train a probabilistic classifier $\hat{P}^{(k)}(d, t)$ to approximate $P^{(k)}(Y = 1|d, t)$ and then set the $k + 1$-th filter in the ensemble to be

$$f^{(k+1)}(d,t) = \frac{1}{2} \log \frac{\hat{P}^{(k)}(d,t)}{(1-\hat{P}^{(k)}(d,t))}. \quad (7)$$

Now, a method commonly used in machine learning to approximate a distribution $q(\xi)$ from a model $\hat{q}(\xi)$ consists in minimizing the so-called cross-entropy loss $J(\hat{q}, q) = -q(\xi)\log\hat{q}(\xi) - (1-q(\xi))\log(1-\hat{q}(\xi))$ on a set of training examples distributed according to $q(\xi)$. We can obtain examples distributed according to our target distribution $P^{(k)}(Y = 1|d, t)$ from the interaction matrix R by defining $R_{ij}^{(k)} = W^{(k)}(d_i, t_j)R_{ij}$. The probabilistic model $\hat{P}^{(k)}(d, t)$ at iteration k can thus be trained to minimize

$$J^{(k)} = -\sum\nolimits_{i,j} R_{ij}^{(k)} \log \hat{P}^{(k)}(d_i, t_j) + (1 - R_{ij}^{(k)}) \log(1 - \hat{P}^{(k)}(d_i, t_j)), \quad (8)$$
$$= -\sum\nolimits_{i,j} W_{ij}^{(k)} \left(R_{ij} \log \hat{P}_{ij}^{(k)} + (1 - R_{ij}) \log(1 - \hat{P}_{ij}^{(k)}) \right),$$

where $W_{ij}^{(k)} = W^{(k)}(d_i, t_j)$ and $\hat{P}_{ij}^{(k)} = \hat{P}^{(k)}(d_i, t_j)$.

Base Learner. Note that if $W_{ij}^{(k)} \propto 1 \, \forall i, j$ Eq. (8) is exactly the objective function employed by NRLMF, except by the constant c, controlling the relative weight of the positive examples, and the regularization terms. That means that a principled method to combine DTI filters correspond to solve re-weighted versions of NRLMF objectives at each iteration, where the weights for each drug-target pair are iteratively updated using Eq. (6). It is hence natural to adopt the probabilistic model used by NRLMF to implement $\hat{P}_{ij}^{(k)}$, that is, set $\hat{P}^{(k)}(d_i, t_j) = \sigma(u_i^k (v_j^k)^T)$ where $\{u_i^{(k)}\}_i, \{v_j^{(k)}\}_j$ correspond to new embeddings for drugs $\{d_i\}_i$ and targets $\{t_j\}_j$. An advantage of this decision is that we can rely on proven methods to fit probabilistic interaction models on DTI data. For instance, we can easily adapt the alternated gradient method employed in [9] to optimize (8). The required derivatives are exactly those employed by standard NRLMF, except that they become scaled by the weight distribution $W_{ij}^{(k)}$. We can also easily incorporate the additional components in the objective functions of NRLMF to each iteration of our boosting procedure in order to handle the high sparsity of the interaction matrices available in typical DTI applications. In practice, the regularization parameters can be tuned using model selection techniques, eventually leading to the plain objective of (8) is this setting is optimal.

Algorithm. The proposed method is summarized as Algorithm 1. As mentioned in the previous paragraph, step 4 of this method can be implemented using alternated gradient descent. Note also that step 3 can be performed recursively exploiting the additivity of $F^{(k)}$.

4 Experiments

We evaluate our method in the Yamanishi dataset collection[1], a gold standard for assessing DTI algorithms. It is composed of 4 prediction problems, namely, enzymes, ion channels (IC), g-protein coupled receptors (GPCR) and nuclear

[1] Yamanishi datasets are publicly available at http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/.

Algorithm 1. Proposed Algorithm (AdaNRLMF).

Initialize the ensemble as $F^{(0)} = 0$;
for $k \leftarrow 0$ **to** $K - 1$ **do**
  Compute the example weights as in Equation (6);
  Implement the model $\hat{P}^{(k)}(d_i, t_j) = \sigma(u_i^k (v_j^k)^T)$ where $\{u_i^{(k)}\}, \{v_j^{(k)}\}$ are new embeddings for drugs $\{d_i\}_i$ and the targets $\{t_j\}_j$ obtained by training NRMLF with example weights $W_{ij}^{(k)}$;
  Set $f^{(k+1)}$ as in Equation (5) and expand the ensemble $F^{(k+1)} = F^{(k)} + f^{(k+1)}$;
end
Return an ensemble of K DTI filters $F^{(K)}$.

receptors (NR) [14], corresponding to different types of proteins. We also consider an updated version of these four datasets, introduced in [7], in which more recently discovered drug-target interactions have been included. As usual, we adopt the *area under precision recall curve* (AUPR) as evaluation metric. This score is preferred over other information retrieval metrics in DTI studies as it illustrates better the differences between algorithms where there are significantly more negative than positive examples.

We compare our method[2] with NRLMF [9], a state-of-the-art method for DTI, and AdaMF [13], an ensemble algorithm for recommendations which is not specialized for DTI. In order to select the optimal parameters for NRLMF, we used a stratified cross-validation scheme. A train/test split is first obtained by randomly selecting 10% of positive and negative interactions for testing and 90% for training. Parameter selection is then performed using 10-fold stratified cross-validation on the resulting training set. That is, the training data is further split into 10 non-overlapping blocks. Each block is retained once as the validation data for evaluating the model that is trained on the remaining 9 blocks. The different results are averaged to produce a single performance estimation. We perform the optimal parameter selection in the same parameter space utilized in NRLMF. In order to speed up the parameter selection process, we adopt the Bayesian optimization method specifically devised for NRLMF in [1]. Once the best parameters for a given train/test split have been determined, the model is trained using the full training set and its output prediction is evaluated in the test set. This stratified cross-validation scheme is repeated 10 times using different train-test splits to obtain more significant results. The performance of the proposed method as well as that of AdaMF are computed on the same train-test splits used for NRMLF. However, in order to select parameters for these methods, we adopt a more simple strategy. For AdaMF, we adopt the parameters suggested by the authors in [13]. For our method, we apply simplifications. First, we train each learner in the ensemble with exactly the same parameters, since it allows to evaluate better the effect of our boosting approach (different

[2] Our code is available at https://gitlab.com/cw_cw/adanrlmf.

parameters may introduce additional diversity in the ensemble not due to the weight distribution adaptation). Second, since we are employing NRMLF as the base learner, we set the base learner parameters to the same values selected for this method in each train-test split. A more exhaustive parameter search may have resulted in slightly better results.

Table 1 shows the average AUPR score obtained by the different methods in each dataset. Standard deviations (computed among the 10 train-test splits) are shown in parenthesis. We can see that the proposed method improves with respect to NRLMF in 7 of 8 datasets, including all the augmented versions of the Yamanishi collection. Our worst result is obtained in the Nuclear Receptors dataset, which correspond to the DTI problem with less known annotations. This may suggest that an ensemble of DTI filters require more positive examples than a single filter to generalize well. Indeed, though the performance of our method in the augmented variants is not always better than the performance observed in the original datasets, the best relative improvements with respect to NRMLF are achieved exactly in those cases, probably because they are more dense in terms of available annotations. Our experiments show also that an ensemble of collaborative filters not specialized for DTI can obtain quite poor results in this type of task. We attribute this result to the fact that AdaMF does not employ information beyond the interaction matrix to predict drug-target interactions, while NRLMF and our method exploit specific knowledge regarding drug and protein similarities to improve their predictions.

Table 1. Average AUPR over 10 trial of 10 fold stratified CV for Yamanishi dataset and its extended version. AdaMF and AdaNRLMF were trained with 10 base learners. Best results are in bold. The last column has the relative improvement of AdaNRLMF with respect to NRLMF.

Dataset	NRLMF	AdaMF	AdaNRLMF	Relative improvement
NR	**0.774 (0.089)**	0.089 (0.025)	0.693 (0.106)	−10.47%
NR Ext.	0.613 (0.099)	0.182 (0.060)	**0.632 (0.082)**	3.09%
GPCR	0.739 (0.073)	0.112 (0.039)	**0.785 (0.035)**	6.22%
GPCR Ext.	0.800 (0.071)	0.538 (0.036)	**0.870 (0.035)**	8.75%
IC	0.899 (0.016)	0.423 (0.038)	**0.943 (0.017)**	4.89%
IC Ext.	0.889 (0.025)	0.562 (0.034)	**0.942 (0.006)**	6.92%
Enzyme	0.881 (0.013)	0.539 (0.030)	**0.909 (0.011)**	3.18%
Enzyme Ext.	0.753 (0.029)	0.488 (0.023)	**0.816 (0.012)**	8.37%

5 Conclusions

In this paper we have devised an Adaboost algorithm specialized for drug-target interaction prediction. It entails solving weighted versions of the objective function underlying NRMLF, a well-known method for this type of problems.

Our experiments show that this method outperforms a single DTI filter and an Adaboost algorithm not specialized for DTI in 7 of 8 datasets. Future work includes the use of the Adaboost as a feature selector, following a multi-kernel approach for DTI. In this variant, several base learners are trained with different similarity measures at each round, and the best predictor is added to the ensemble.

Acknowledgements. This research was partially supported by PIIC-2018 program of DGIP from the Federico Santa María Technical University.

References

1. Ban, T., Ohue, M., Akiyama, Y.: Efficient hyperparameter optimization by using Bayesian optimization for drug-target interaction prediction. In: IEEE 7th ICCABS, pp. 1–6, October 2017
2. Cobanoglu, M.C., Liu, C., Hu, F., Oltvai, Z.N., Bahar, I.: Predicting drug-target interactions using probabilistic matrix factorization. J. Chem. Inf. Model. **53**(12), 3399–3409 (2013)
3. Ding, H., Takigawa, I., Mamitsuka, H., Zhu, S.: Similarity-based machine learning methods for predicting drug-target interactions: a brief review. Briefings Bioinform. **15**(5), 734 (2014). https://doi.org/10.1093/bib/bbt056
4. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Ann. Stat. **38**(2), 337–407 (2000)
5. Gönen, M.: Predicting drug-target interactions from chemical and genomic kernels using bayesian matrix factorization. Bioinformatics **28**(18), 2304–2310 (2012)
6. Johnson, C.C.: Logistic matrix factorization for implicit feedback data. In: Advances in Neural Information Processing Systems 27 (2014)
7. Keum, J., Nam, H.: Self-BLM: prediction of drug-target interactions via self-training SVM. PLOS ONE **12**(2), 1–16 (2017). https://doi.org/10.1371/journal.pone.0171839
8. Li, Z., et al.: In silico prediction of drug-target interaction networks based on drug chemical structure and protein sequences. Sci. Rep. **7**(1), 11174 (2017)
9. Liu, Y., Wu, M., Miao, C., Zhao, P., Li, X.L.: Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. PLOS Comput. Biol. **12**(2), 1–26 (2016). https://doi.org/10.1371/journal.pcbi.1004760
10. Rayhan, F., Ahmed, S., Shatabda, S., Farid, D.M., Mousavian, Z., Dehzangi, A., Rahman, M.S.: iDTI-ESBoost: identification of drug target interaction using evolutionary and structural features with boosting. Sci. Rep. **7**(1), 17731 (2017)
11. Smola, A.J., Schölkopf, B.: Learning with kernels, vol. 4. Citeseer (1998)
12. Tsai, C.F., Hung, C.: Cluster ensembles in collaborative filtering recommendation. Appl. Soft Comput. **12**(4), 1417–1425 (2012)
13. Wang, Y., Sun, H., Zhang, R.: AdaMF: adaptive boosting matrix factorization for recommender system. In: Li, F., Li, G., Hwang, S., Yao, B., Zhang, Z. (eds.) WAIM 2014. LNCS, vol. 8485, pp. 43–54. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08010-9_7
14. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., Kanehisa, M.: Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics **24**(13), i232 (2008). https://doi.org/10.1093/bioinformatics/btn162

15. Yuan, Q., Gao, J., Wu, D., Zhang, S., Mamitsuka, H., Zhu, S.: Druge-rank: improving drugtarget interaction prediction of new candidate drugs or targets by ensemble learning to rank. Bioinformatics **32**(12), i18–i27 (2016)
16. Zheng, X., Ding, H., Mamitsuka, H., Zhu, S.: Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. In: Proceedings of the 19th ACM SIGKDD, pp. 1025–1033. ACM (2013)