

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Predicting Drug Targets from Heterogeneous Spaces using Anchor Graph Hashing and Ensemble Learning

Yi Zheng*, Hui Peng*, Xiaocai Zhang*, Xiaoying Gao[†], and Jinyan Li*

*Advanced Analytics Institute, University of Technology Sydney, Sydney 2007, Australia.

[†]School of Engineering and Computer Science, Victoria University of Wellington, Wellington 6140, New Zealand.

Emails: Jinyan.Li@uts.edu.au (Jinyan Li) and Yi.Zheng-8@student.uts.edu.au (Yi Zheng)

Abstract—The *in silico* prediction of potential drug-target interactions is of critical importance in drug research. Existing computational methods have achieved remarkable prediction accuracy, however usually obtain poor prediction efficiency due to computational problems. To improve the prediction efficiency, we propose to predict drug targets based on integration of heterogeneous features with anchor graph hashing and ensemble learning. First, we encode each drug as a 5682-bit vector, and each target as a 4198-bit vector using their heterogeneous features respectively. Then, these vectors are embedded into low-dimensional Hamming Space using anchor graph hashing. Next, we append hashing bits of a target to hashing bits of a drug as a vector to represent the drug-target pair. Finally, vectors of positive samples composed of known drug-target pairs and randomly selected negative samples are used to train and evaluate the ensemble learning model. The performance of the proposed method is evaluated on simulative target prediction of 1094 drugs from DrugBank. Extensive comparison experiments demonstrate that the proposed method can achieve high prediction efficiency while preserving satisfactory accuracy. In fact, it is 99.3 times faster and only 0.001 less in AUC than the best literature method “Pairwise Kernel Method”.

Index Terms—drug target prediction; heterogeneous features; anchor graph hashing; ensemble learning.

I. INTRODUCTION

The identification of interactions between drugs and targets is a key element in drug discovery. It helps to identify new drugs for known targets or novel targets for existing drugs. Though various biological assays are available nowadays, experimental identification of drug-target interaction remains challenging [1]. To guide and speed up the expensive and laborious experimental approaches, a variety of *in silico* prediction methods have been developed [2]. These approaches can be classified into two types, namely docking simulation and machine learning. Docking simulation is a powerful approach, however, it requires 3D structure information of the target proteins [3]. Moreover, it is time-consuming, in that a large amount of computational resources are required [4]. Therefore, it’s hard to apply this technique on a large scale. By comparison, machine learning is more efficient and capable in large-scale predictions.

Machine learning methods are based on the observation that similar drugs tend to interact with similar targets [2]. The prediction is performed based on information such as

drug chemical structures [5] [6], drug side-effects [7], target protein sequences [8], and the known drug-target interactions. A straightforward approach is to use binary classification methods where the drug-target pairs are vectorized as input for binary classifiers, e.g., artificial neural network (ANN) and support vector machine (SVM) [6] [8] [9] [10]. Another encouraging approach named bipartite local model (BLM), is to view drugs and targets as a bipartite graph. BLM builds one classifier from the drug side and one classifier from the target side for each drug-target interaction to be predicted. It can achieve remarkable accuracy, however, leads to a serious computational problem because a large number of classifiers have to be trained. The current state-of-the-art for predicting drug-target interactions using machine learning is to embed drug similarity measures and target similarity measures into kernels [6] [8] [11] [12]. By using kernels, multiple information sources can be integrated to achieve high prediction accuracy. However, one serious problem is the scale of the training set with “the number of training drugs times the number of training targets”, bringing computational difficulties in large-scale applications.

As described above, existing machine learning methods either use vectors extracted from drug-target pairs, or the similarity matrix as input for predictions. The dimension of the input vectors and the similarity matrix is quite high, which usually leads to serious computational problems. To tackle this problem, we introduce anchor graph hashing (AGH), an excellent hashing method in both the speed and accuracy [13]. It can embed the data into Hamming space so that the neighbors in the original data space remain neighbors in the Hamming space [14]. Therefore, AGH is an ideal technique for dimension reduction. Recently, ensemble learning has been widely used in bioinformatics [15]. It combines multiple hypotheses of the base learners to form a better hypothesis, thus producing better predictive performance. Lots of application have proved the success of ensemble learning [16] [17] [18].

In this work, we propose to predict drug-target interactions from heterogeneous spaces with anchor graph hashing and ensemble learning (AGHEL). First, every drug is encoded as a vector using its heterogeneous features including drug chemical structures, drug side-effects, and drug substituents. Analogously, each target is encoded as a vector using its

related Gene Ontology (GO). Next, the high-dimensional drug and target vectors are compressed into short vectors (i.e., hashing bits) in Hamming space using AGH. Then we can get the compressed vector of arbitrary drug-target pair by appending the compressed target vector to the compressed drug vector. Finally, compressed vectors of drug-target pairs are used as input to train and evaluate the ensemble learning model. We employ random forest [19] and XGBoost [20] as base learners of the ensemble learning model. Our innovation lies in integrating heterogeneous drug and target features into a unified framework for predictions, in utilizing the advanced anchor graph hashing to reduce vector dimension, and in the employment of ensemble learning to improve the prediction performance.

II. RELATED WORK

A plenty of prediction models have been developed to predict drug-target interactions. We discuss three typical models which we use as comparison methods, namely nearest neighbor (NN) algorithm, pairwise kernel method (PKM), and bipartite local models (BLM).

A. Nearest Neighbor

NN is usually used as the baseline method [8]. Given a test candidate drug, NN predicts it to interact with targets known to interact with its nearest drug. Likewise, NN predicts a test candidate target to interact with drugs known to interact with its nearest target. Then each interaction has two prediction results/scores which will be further combined to give a definitive prediction. NN is simple and efficient, however, the performance is relatively low.

B. Pairwise Kernel Method

PKM is a typical kernel based method, which uses drug-target interactions as instances [4]. Given two drug pairs (d, t) and (d', t') , PKM first computes the similarity between them by multiplying the drug similarity score and the target similarity score as follows:

$$s((d, t), (d', t')) = s_d(d, d') * s_t(t, t') \quad (1)$$

where s , s_d , and s_t are similarity functions of interactions, drugs and targets respectively, d and d' are drugs, and t and t' are targets. Then it uses the calculated similarity matrix of drug-target pairs as the self-defined kernel to train an SVM classifier. Finally, the trained SVM predicts scores for test drug-target pairs. By using kernels, PKM can integrate multiple information sources and achieve high prediction accuracy. However, it leads to computational difficulties in large-scale problems due to the scale of training set. For instance, for the 1094 drugs and 1556 targets used in this work, the size of the kernel matrix will be $(1094*1556)*(1094*1556)$ which is quite large.

C. Bipartite Local Models

BLM views drug-target interactions as a bipartite network, in which vertexes are drugs and targets, edges are their interactions. Each drug-target interaction is predicted from the drug side and target side respectively using a local model (classifier). We take the prediction of a drug-target pair (d, t) as an example to illustrate the process. BLM first checks whether every target in the training set known to interact with drug d , and assign a label "1" if so; otherwise "0"; Next, vectors of all training targets or their similarity matrix together with their labels are used to train a local model. Then the trained local model predicts the label of t . Analogously, BLM trains a local model using drugs in the training set and obtain the label of d . Finally, two prediction scores of d and t are combined to give a final prediction. BLM builds one classifier from the drug side and one classifier from the target side for each drug-target pair. It can achieve remarkable accuracy. However, BLM leads to a serious computational problem because a large number of classifiers (i.e., twice of interactions to be predicted) have to be trained.

III. DRUG TARGET PREDICTION WITH ANCHOR GRAPH HASHING AND ENSEMBLE LEARNING

To tackle the computational problem of large-scale drug target predictions while keeping the prediction accuracy, we introduce two advanced techniques namely anchor graph hashing and ensemble learning for predictions. Anchor graph hashing is capable to embed high-dimensional data into low-dimensional Hamming space while keeping neighbors in the original data space remaining neighbors in Hamming space. That's why we employ it to reduce the dimension of raw drug vectors and target vectors. Ensemble learning has shown its advantages in many applications of bioinformatics. We make full use of its advantage to better capture the data features and obtain better performance.

A. Prediction Framework

The framework of our method is illustrated in Figure 1. First, each drug is encoded as a 5682-bit vector using its chemical substructures (881), side-effects (4063), and substituents (738) where each bit represents the presence or absence of the feature by 1 or 0. Likewise, every target is encoded as a vector of 4,198 bits using the Gene Ontology. Then, we employ AGH to embed raw drug and target vectors into low-dimensional Hamming Space (12/16/24/32/48/64 bits). After that, we use all known drug-target interactions as positive samples, and randomly select the same number of negative samples from unknown interactions. Next, sample vectors formed by appending corresponding target hashing bits to the drug hashing bits are used as input to train and evaluate the ensemble learning model. The ensemble learning model employs random forest and XGBoost as base learners, and gives the final prediction scores by averaging the scores predicted by them. Finally, the prediction results are evaluated through 10-fold cross-validation on both the prediction accuracy and prediction efficiency.

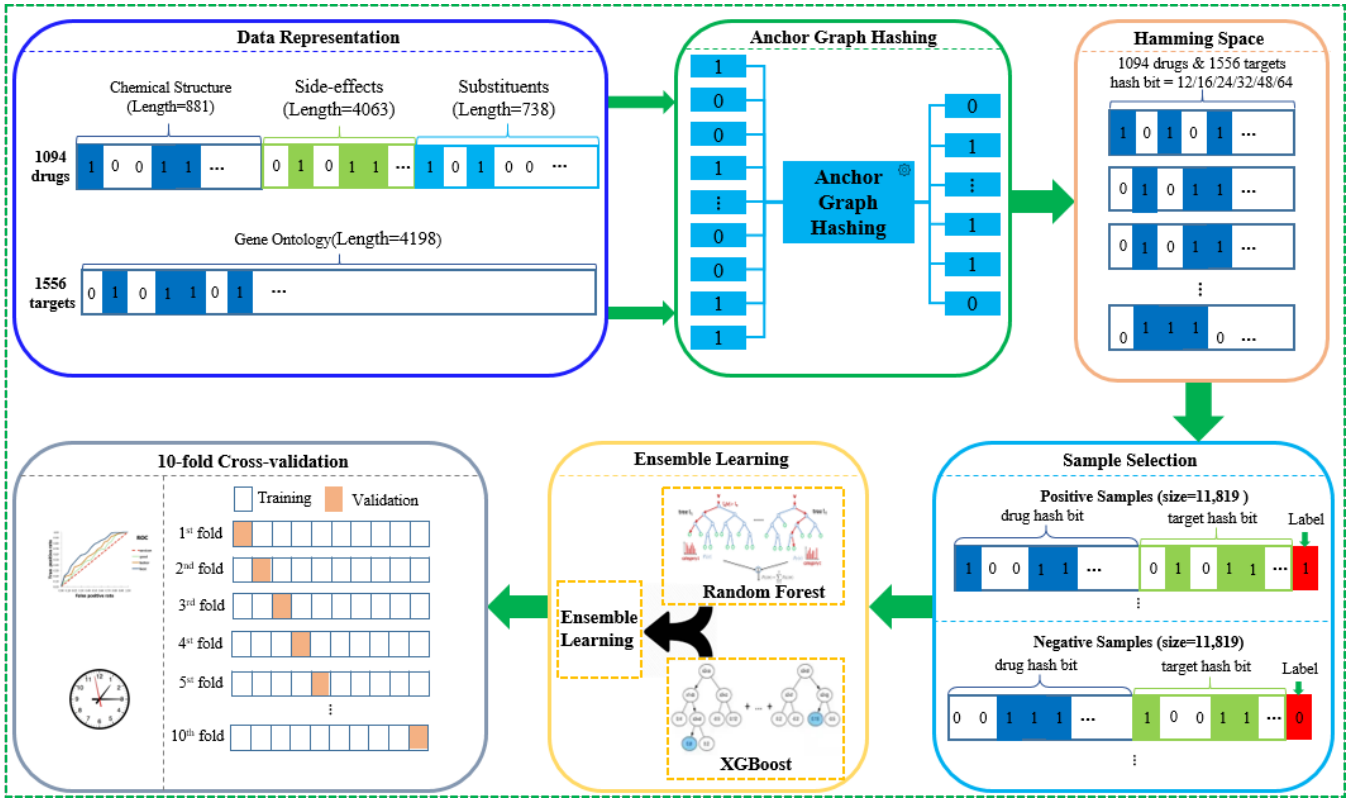


Fig. 1. The framework for drug target prediction by anchor graph hashing and ensemble learning.

B. Data Representation

We collected 1094 drugs and 1556 targets for experiments. Each drug was represented by a 5682 dimensional binary vector using its chemical structure (881), associated side-effects (4063), and substituents (738). The drug chemical structure is represented by 881 chemical substructures defined in PubMed [21]. We obtained 4063 associated side-effects from SIDER [22] and 738 substituents from DrugBank [23]. The elements of the drug vector encode for the presence or absence of each of the features (i.e., chemical substructures/side-effects/substituents) by 1 or 0. We obtained 4198 unique GO terms which are associated with the targets from the EMBL-EBI website [24]. Analogously, we represented each target as a 4198 dimensional binary vector where each element encodes the presence or absence of each of the GO terms by 1 or 0.

C. Anchor Graph Hashing

The goal of AGH is to learn short binary codes such that neighbors in the input space are mapped to neighbors in the Hamming space [14]. AGH seeks an r -bit Hamming embedding $Y \in \{1, -1\}^{n \times r}$ for n points in the input space by minimizing

$$\begin{aligned} \min_Y \frac{1}{2} \sum_{i,j=1}^n \|Y_i - Y_j\|^2 A_{ij} &= \text{tr}(Y^\top LY) \\ \text{s.t. } Y \in \{1, -1\}^{n \times r}, 1^\top Y &= 0, Y^\top Y = nI_{r \times r} \end{aligned} \quad (2)$$

where Y_j is the j^{th} row of Y representing the r -bit code for point x_j , and A_{ij} is the similarity between the data pair (x_i, x_j) . To solve the above NP-hard problem, AGH uses a small subset of m points called anchors to approximate the data neighborhood structure [25]. Specifically, K-means is performed on all n points to get m ($m \ll n$) cluster centers (anchors) $U = \{u_j \in \mathbb{R}^d\}_{j=1}^m$ first. Next AGH computes truncated similarity matrix between all n data points and m anchors as follows:

$$Z_{ij} = \begin{cases} \frac{\exp(-D^2(x_i, u_j)/t)}{\sum_{j' \in \langle i \rangle} \exp(-D^2(x_i, u_{j'})/t)} & , \forall j \in \langle i \rangle \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $\langle i \rangle \in [1 : m]$ denotes the indices of s ($s < m$) nearest anchors of point x_i in U according to a distance function $D(\cdot)$, and t is the bandwidth parameter. Then the adjacent matrix A can be approximated as $A' = ZV^{-1}Z^\top$ where $V = \text{diag}(Z^\top \mathbf{1}) \in \mathbb{R}^{m \times m}$ [25]. The eigenvectors of A' can be calculated easily by using its low-rank property. In particular, AGH solves the eigenvalue system of a small $m \times m$ matrix $M = V^{-1/2}Z^\top ZV^{-1/2}$. This produces r eigenvector-eigenvalue pairs $\{(p_k, \sigma_k)\}_{k=1}^r$ where $0 < \sigma_r \leq \dots \leq \sigma_1 < 1$. Finally, AGH obtains the desired spectral embedding matrix as

$$Y = \sqrt{n}ZV^{-1/2}P \sum^{-1/2} \quad (4)$$

where $P = [p_1, \dots, p_r] \in \mathbb{R}^{m \times r}$ and $\sum = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r \times r}$.

Using AGH, the high-dimension input vectors are compressed into low-dimension hashing bits where neighbours in the input space remain neighbours in the output space.

D. Sample Selection

All known drug-target pairs were used as positive samples and assigned the label “1”. An equal number of negative samples were randomly selected from unknown drug-target pairs and assigned the label “0”. To avoid bias, the negative sample selection process were repeated 5 times and the average results were used for evaluation. After we obtained positive samples and negative samples, we represented them as vectors by appending corresponding target hashing bits to the drug hashing bits. Then these vectors were used as input of the ensemble learning model.

E. Ensemble Learning

We chose random forest and XGBoost as the base learners of the ensemble learning model as they perform well on low-dimension data. We simply took the average value of their prediction scores as the final prediction score.

F. 10-fold Cross-validation

The performance is evaluated by 10-fold cross validation: (1) samples in the gold standard are split into 10 roughly equal-sized subsets; (2) each subset is taken in turn as the test set, and the remaining subsets are used as training set; (3) all results over the 10-fold validation are used for evaluation. We randomly chose the same number of negative samples as that of positive samples (known drug-target interactions). To avoid bias, we repeated the sample selection process and cross-validation 5 times respectively, which means each method would run 25 (5*5) times and the average results were used for final evaluation. We employed the receiver operating characteristic curve (ROC), the area under ROC curve (AUC) and, the execution time as the evaluation metrics to evaluate the prediction performance.

IV. RESULTS

A. Data Resources

Since drug-target interactions are continually being identified with time. Instead of using old data from existing literature, we collected the dataset from well-known databases by ourselves. Specifically, the drugs were collected from DrugBank [23], a comprehensive drug database. The targets were obtained from DrugBank and DrugCentral [26]. Table I and Figure 2 show the details of our dataset (Dataset_AGHEL). By comparison, we also list the datasets used in PKM (Dataset_PKM) and BLM (Dataset_BLM). From table I, we can see that our dataset has much more targets and know interactions.

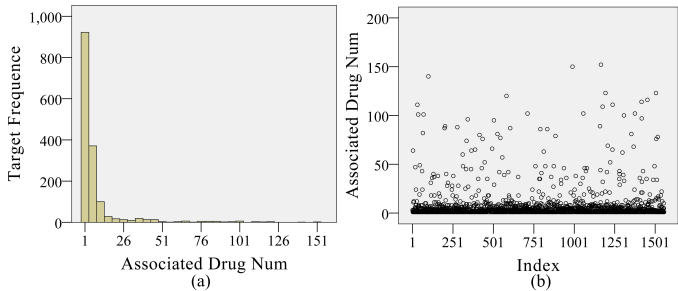


Fig. 2. Characteristics of targets and their associated drugs. The left panel is the histogram of the associated drug number for the targets and the right panel is the index-plot of the number of associated drugs for each target.

TABLE I
DATASET USED IN THIS WORK AND TWO DATASETS USED IN PUBLICATIONS.

Dataset	Number of drugs	Number of targets	Number of interactions
Dataset_AGHEL	1094	1556	11,819
Dataset_BLM	932	989	5127
Dataset_PKM	1205	889	2782

B. Parameter Optimization

The key parameters of anchor graph hashing are the number of its layer and the output hashing bits [14]. Therefore, we performed experiments to find out the best settings for the hashing layer and output hashing bit. Specifically, we employed K-means to cluster all raw drug (target) vectors to get 100 cluster centers as their anchors first. Then we obtained r -bit ($r \in \{12, 16, 24, 32, 48, 64\}$) hashing bits for drugs(targets) with their anchors and AGH of different number of layers. Finally, the drug hashing bits and the target hashing bits obtained from the same number of layers and bits are combined as vectors for predictions and evaluations. The experiment results are illustrated in Figure 3. It can be seen obviously that the AUC increases with the number of hashing bits for both 1-layer AGH and 2-layer AGH. Besides, 1-layer AGH outperforms 2-layer AGH on the same number of bits. Based on the above observation, we set the layer of AGH as 1 and the number of output hashing bits as 64 for AGHEL in the following experiment.

C. Comparison Results

To demonstrate the advantage of our method, we compared AGHEL with three different types of existing prediction methods on both the prediction accuracy and prediction efficiency. Details of the three comparison methods namely NN, PKM, and BLM, are detailedly introduced in the “Related Work” section. All methods were performed on a Linux workstation with dual Intel XEON 3.4GHz processors (14 cores) and 256 GB of main memory.

Figure 4 shows the average execution time of the four methods. We can see that AGHEL is the fastest, completed predictions of 23,638 drug-target interactions in 9.99 s (0.53s for hashing compression and 9.46s for predictions). It is 4.62

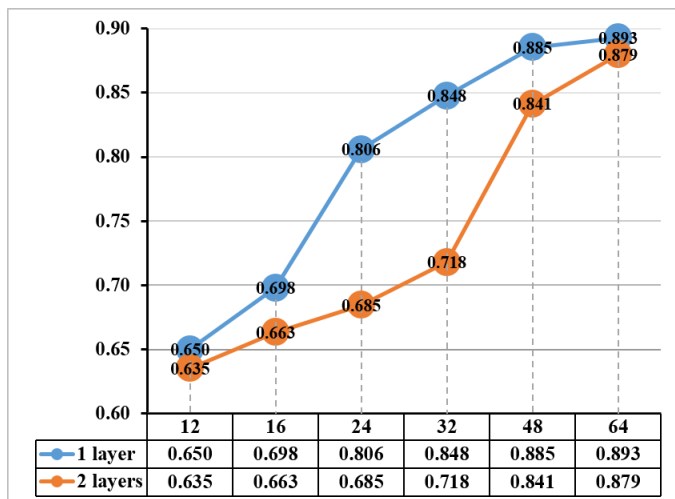


Fig. 3. The AUC (ROC score) of AGHEL with different AGH settings. The x-axis is the number of output hashing bits and the y-axis is the AUC score.

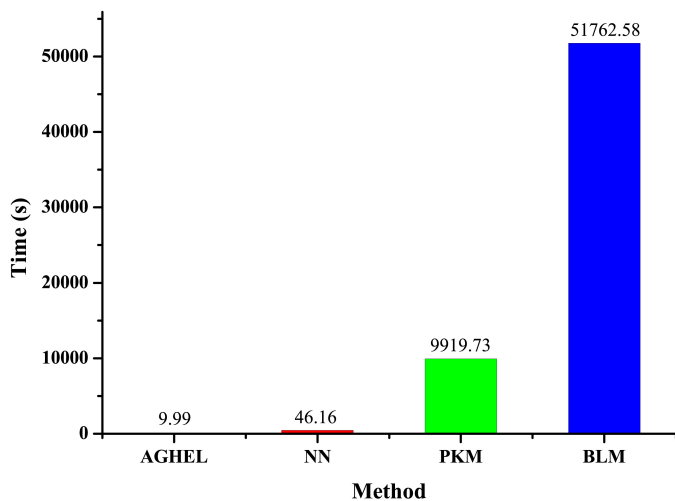


Fig. 4. The average execution time (s) of four methods using 10-fold cross-validation.

times faster than the second fastest method NN, and 5181.44 times faster than the slowest method BLM. Among the three existing comparison methods, NN is the fastest, benefits from its simple prediction principle i.e., predicting the interacted targets (drugs) for drugs(targets) the same as its nearest drug (target). BLM is the slowest because two classifiers need to be trained for each interactions. It means 47,276 classifiers have to be trained for predicting the whole 23,638 interactions. By comparison, PKM is more efficient than BLM, as only one classifier needs to be trained. However due to the large size of the kernel matrix ($23,638 \times 23,638$), it is still much slower than NN and AGHEL. AGHEL trains one classifier for all predictions as well. However, compared with PKM whose input vector dimension equals to 23,638, the input vector dimension of AGHEL is largely reduced from 9,880 (5,682+4,198) to 128 (64+64) by its AGH component. Through this fast process (0.53s), the prediction time of AGHEL decreases largely. This

explains why AGHEL costs the least execution time.

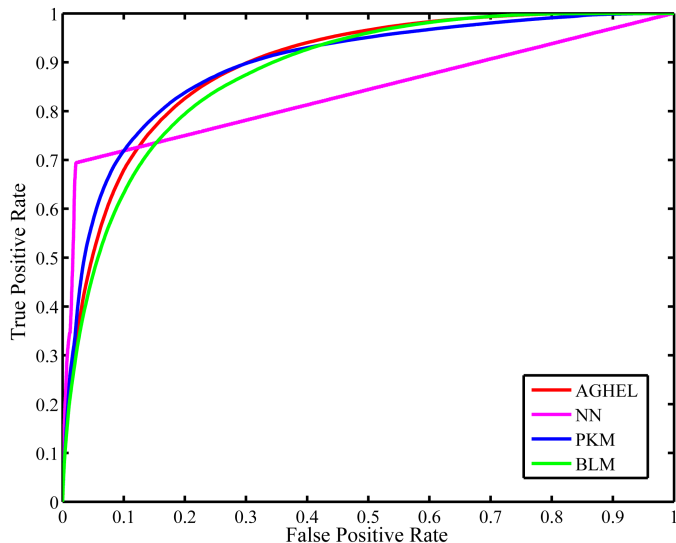


Fig. 5. The ROC curve of four methods from 10-fold cross-validation experiments.

To draw the ROC curves, we merged all prediction scores from the 25 repeated experiments for each method and a single ROC curve was drawn. Figure 5 illustrates the obtained ROC curves. From it, we can see that PKM (colored blue) achieves the best performance. It has been demonstrated to be the most powerful method in one review of machine learning methods for drug target predictions [4]. The proposed method (colored red) AGHEL achieves the second best performance. To better visualize the differences between the four methods, we computed the average AUC for each method by averaging AUCs from its 25 repeated experiments. The average AUCs are shown in Figure 6. It can be seen that AGHEL (0.893) achieves satisfactory AUC, only 0.001 less than PKM (0.894). Following AGHEL are BLM (0.879) and NN (0.836).

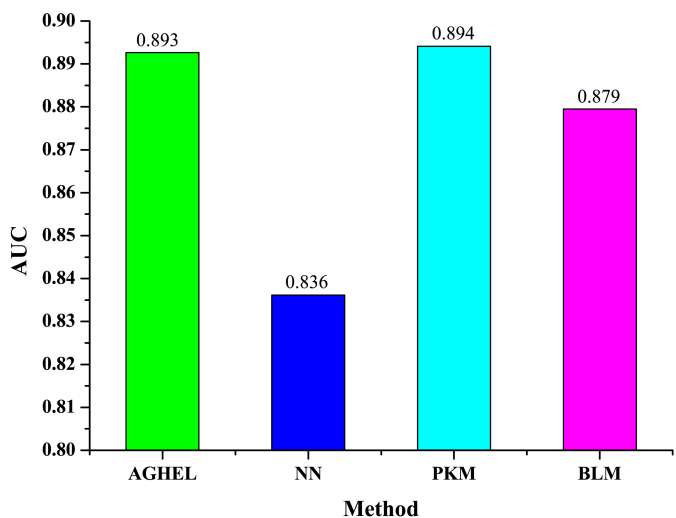


Fig. 6. The AUC of four methods from 10-fold cross-validation experiments.

V. CONCLUSION

In this work, we propose to predict drug targets based on integration of heterogeneous features with anchor graph hashing and ensemble learning. The contribution of the proposed method lies in the integration of drug chemical structures, drug side-effects, drug substituents and GO terms of targets in a unified framework for predictions, in the novel use of anchor graph hashing, and in the implement of ensemble learning. To our knowledge, no existing work gathers the above features in the field of drug target prediction. From the experimental results, it has been demonstrated that the proposed method, namely AGHEL achieves satisfactory performance on both the execution time and prediction accuracy. Compared with three state-of-the-art methods, it is the fastest and the second most accurate. Owing to the introduction of anchor graph hashing, the prediction process is sharply sped up by decreasing dimensions of input vectors. With the help of integration of heterogeneous features of drugs and targets, and the employment of ensemble learning, AGHEL achieves high prediction accuracy.

The proposed method is of value to the drug discovery process, which aims to identify novel targets for existing drugs or new drugs for known targets. It can guide and speed up the laborious, expensive and tedious experimental identification of drug-target interaction [1]. In this work, drug chemical structures, drug side-effects, drug substituents and GO terms of targets are integrated to predict drug targets. It should be pointed out that more drug and target information, such as drug indications and target sequences, can be integrated. The limitation of the proposed method is its dependence on the used drug and target features which are not always available and complete for all drugs. Therefore, to develop methods which are able to predict drug targets efficiently and accurately with easy accessible data will be our future work.

REFERENCES

- [1] S. J. Haggarty, K. M. Koeller, J. C. Wong, R. A. Butcher, and S. L. Schreiber, "Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays," *Chemistry & biology*, vol. 10, no. 5, pp. 383–396, 2003.
- [2] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori, "Gaussian interaction profile kernels for predicting drug–target interaction," *Bioinformatics*, vol. 27, no. 21, pp. 3036–3043, 2011.
- [3] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov, "Principles of docking: An overview of search algorithms and a guide to scoring functions," *Proteins: Structure, Function, and Bioinformatics*, vol. 47, no. 4, pp. 409–443, 2002.
- [4] H. Ding, I. Takigawa, H. Mamitsuka, and S. Zhu, "Similarity-based machine learning methods for predicting drug–target interactions: a brief review," *Briefings in Bioinformatics*, vol. 15, no. 5, pp. 734–747, 2013.
- [5] Y. C. Martin, J. L. Kofron, and L. M. Traphagen, "Do structurally similar molecules have similar biological activity?" *Journal of medicinal chemistry*, vol. 45, no. 19, pp. 4350–4358, 2002.
- [6] N. Nagamine and Y. Sakakibara, "Statistical prediction of protein–chemical interactions based on chemical structure and mass spectrometry data," *Bioinformatics*, vol. 23, no. 15, pp. 2004–2012, 2007.
- [7] 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.
- [8] K. Bleakley and Y. Yamanishi, "Supervised prediction of drug–target interactions using bipartite local models," *Bioinformatics*, vol. 25, no. 18, pp. 2397–2403, 2009.
- [9] J. R. Bock and D. A. Gough, "Virtual screen for ligands of orphan g protein-coupled receptors," *Journal of chemical information and modeling*, vol. 45, no. 5, pp. 1402–1414, 2005.
- [10] J.-L. Faulon, M. Misra, S. Martin, K. Sale, and R. Sapra, "Genome scale enzyme–metabolite and drug–target interaction predictions using the signature molecular descriptor," *Bioinformatics*, vol. 24, no. 2, pp. 225–233, 2007.
- [11] L. Jacob and J.-P. Vert, "Protein-ligand interaction prediction: an improved chemogenomics approach," *Bioinformatics*, vol. 24, no. 19, pp. 2149–2156, 2008.
- [12] A. M. Wassermann, H. Geppert, and J. Bajorath, "Ligand prediction for orphan targets using support vector machines and various target-ligand kernels is dominated by nearest neighbor effects," *Journal of chemical information and modeling*, vol. 49, no. 10, pp. 2155–2167, 2009.
- [13] Y. Zheng, C. Sun, C. Zhu, X. Lan, X. Fu, and W. Han, "Lwcs: A large-scale web page classification system based on anchor graph hashing," in *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2015, pp. 90–94.
- [14] W. Liu, J. Wang, S. Kumar, and S.-F. Chang, "Hashing with graphs," in *Proceedings of the 28th international conference on machine learning (ICML-11)*. Citeseer, 2011, pp. 1–8.
- [15] Z.-H. Zhou, "Ensemble learning," *Encyclopedia of biometrics*, pp. 411–416, 2015.
- [16] C. Wong, Y. Li, C. Lee, and C.-H. Huang, "Ensemble learning algorithms for classification of mtDNA into haplogroups," *Briefings in bioinformatics*, vol. 12, no. 1, pp. 1–9, 2010.
- [17] P. Yang, P. D. Yoo, J. Fernando, B. B. Zhou, Z. Zhang, and A. Y. Zomaya, "Sample subset optimization techniques for imbalanced and ensemble learning problems in bioinformatics applications," *IEEE transactions on cybernetics*, vol. 44, no. 3, pp. 445–455, 2014.
- [18] W. Zhang, Y. Niu, Y. Xiong, M. Zhao, R. Yu, and J. Liu, "Computational prediction of conformational B-cell epitopes from antigen primary structures by ensemble learning," *PloS one*, vol. 7, no. 8, p. e43575, 2012.
- [19] C. Tomasi, "Random forest classifiers," 2017.
- [20] T. Chen and C. Guestrin, "Xgboost: A scalable tree

- boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.
- [21] B. Chen, D. Wild, and R. Guha, “Pubchem as a source of polypharmacology,” *Journal of chemical information and modeling*, vol. 49, no. 9, pp. 2044–2055, 2009.
- [22] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, “The sider database of drugs and side effects,” *Nucleic acids research*, p. gkv1075, 2015.
- [23] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu *et al.*, “Drugbank 4.0: shedding new light on drug metabolism,” *Nucleic acids research*, vol. 42, no. D1, pp. D1091–D1097, 2013.
- [24] G. O. Consortium, “The gene ontology (go) database and informatics resource,” *Nucleic acids research*, vol. 32, no. suppl 1, pp. D258–D261, 2004.
- [25] W. Liu, J. He, and S.-F. Chang, “Large graph construction for scalable semi-supervised learning,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 679–686.
- [26] O. Ursu, J. Holmes, J. Knockel, C. G. Bologna, J. J. Yang, S. L. Mathias, S. J. Nelson, and T. I. Oprea, “Drugcentral: online drug compendium,” *Nucleic acids research*, vol. 45, no. D1, pp. D932–D939, 2017.