Exploring a Siamese Neural Network Architecture for Drug Discovery

Luis Torres
uc2015241578@student.uc.pt
Department of Informatics Engineering, University of Coimbra

Bernardete Ribeiro
bribeiro@dei.uc.pt
Department of Informatics Engineering, University of Coimbra

Joel Arrais
jpa@dei.uc.pt
Department of Informatics Engineering, University of Coimbra

Abstract

Deep neural networks offer a great predictive power when inferring the pharmacological properties and biological activities of small molecules in drug discovery applications. However, in the traditional drug discovery process, where supervised data is scarce, the lead-optimization step is a low-data problem, making it difficult to find molecules with the desired therapeutic activity and obtain accurate predictions for candidate compounds. One major requirement to ensure the validity of the obtained neural network models is the need for a large number of training examples per class, which is not always feasible in drug discovery applications. This invalidates the use of instances whose classes were not considered in the training phase or in data where the number of classes is high and oscillates dynamically.

The main objective of the study is to optimize the discovery of novel compounds based on a reduced set of candidate drugs. We propose a Siamese neural network architecture for one-shot classification, based on Convolutional Neural Networks (CNNs), that learns from a similarity score between two input molecules according to a given similarity function.

Using a one-shot learning strategy, few instances per class are needed for training, and a small amount of data and computational resources are required to build an accurate model. The results achieved demonstrate that using a Siamese Deep Neural Network for one-shot classification leads to overall improved performance when compared to other state-of-the-art models. The proposed architecture provides an accurate and reliable prediction of novel compounds considering the lack of biological data available for drug discovery tasks.

Introduction

In drug discovery, we seek to maintain the desired properties of the main components of the molecules, preventing any structural deviation that might compromise their biological activity. Thus, the main objective is to discover novel compounds with optimal therapeutic effects, less toxicity, greater pharmacological activity, reduced risks for the organism, and better conditions of solubility and selectivity for the candidate molecules [1].

The feasibility of recognizing new compounds and their pharmacological analogs with a reduced set of biological data available for training remains an important challenge in compound prediction for drug discovery applications. Moreover, the identification of the class whenever a new group of molecules is observed, without requiring a periodic retraining and using only a few training examples per class, is crucial in drug discovery tasks.

Humans are able to learn multiple representations from a small number of examples, and then use the knowledge acquired to distinguishing new examples of these same representations, even if observed only once [2]. These idea of similarity gave rise to one-shot learning methods.

Instead of directly classifying a given instance, a one-shot learning model learns a similarity function that accepts two inputs, and returns a score that denotes the similarity between them. The learnt similarity rule allows to predict instances whose classes are unknown at training stage. The model learns a distance metric capable of distinguish two different inputs, and highlight the dissimilarities between them [3].

In the context of drug discovery, the application of a one-shot classification strategy improves the prediction of novel compounds whose classes are less-represented and only requires one example per class for training. Despite the size of the training set, a single molecule per class is needed for training. This molecule is used as a reference instance to compute the distance with any other molecule, while predicting a novel compound in one shot, according to the output similarity score generated between them. This similarity measure is the probability of both inputs belonging to the same class of molecular structures.

Model

A Siamese neural network built upon two parallel and identical convolutional neural networks is introduced as the proposed model approach. This network is compatible with a set of pairs of compounds provided for training. The model learns a similarity function and returns a distance metric applied to the output feature vectors from both siamese twins. This similarity measure allows the model to predict novel compounds in one shot, based on a reduced set of candidate molecules available for training.

Tox21 was the dataset used to extract SMILES (Simplified Molecular Input Line Entry System) for compound representation and encoding [4].

One-Shot Siamese Neural Network

We propose a Siamese Neural Network that accepts molecules organized in pairs. This model consists in two parallel and identical convolutional neural networks. Both Siamese twins are indistinguishable, since they are two copies of the same network and share the same set of parameters [5].

These parallel networks reduce their respective inputs to increasingly smaller tensors as we progress to the high-level layers. The difference between the output feature vectors is used as an input to the learnt similarity function. In a one-shot learning approach, one compound is established as a reference molecule and compared with different compounds expressing the probability of both belonging to the same class, according to a given similarity score *score*. The Siamese twins are symmetric neural networks, which means that the similarity score generated between d_1 and d_2 is equal to the score generated between d_2 and d_1. Thus, if we switch the order of the inputs of the Siamese network the returned output prediction would be the same:

$$score\,(d_1, d_2) = score\,(d_2, d_1) \tag{1}$$

This symmetry property is very important when learning a similarity metric. An architecture based on two parallel neural networks propagates two inputs through the same set of weights and the difference between the output feature vectors serves as an input to a similarity metric. This symmetry-based approach is less expensive and leads to a pairwise training which improves the model prediction accuracy.

Pairwise Training

A training set in which half are pairs of the same class and another half of different classes was considered. Since the Siamese neural network accepts pairs of molecules, the dataset size increases, given the number of possible combinations for the pairs of molecules available for training. However, we consider half of the pairs of the same class and half of the different classes for training. Therefore, the maximum number of possible combinations for compound pairs is the total number of possible pairs for compounds of the same class. If there are L examples each of Q classes, the total number of possible pairs of the same class is given by,

$$number\ of\ pairs = L \cdot \binom{Q}{2} = \frac{L \cdot Q!}{2! \cdot (Q-2)!} \tag{2}$$

The number of training instances increases in Q of a square factor and in L of a linear factor. The increase in the size of the training set reduces the effect of overfitting.

Model Architecture

The model architecture that maximizes performance is the one whose number of convolutional layers is 4, the number of filters in each layer is a multiple of 16, and in which the corresponding output features maps are applied to a ReLu activation function and to a maxpooling layer.

The output feature map of the last convolutional layer is flattened into a single vector that serves as an input to a fully connected layer with 1024 units. This layer learns a similarity function between two feature vectors by applying a distance metric to the learned feature map. It is followed by a dense layer that computes the absolute difference between the two output feature vectors. This value serves as input to a sigmoid function in the last layer.The predicted similarity score is given by,

$$score = sigmoid\ (\sum_i |\ v_{1,l}^i - v_{2,l}^i\ |) \tag{3}$$

v_1 and v_2 are the output feature vectors of the last convolutional layer of each Siamese twin, l the index representing the dense layer, i the index in each output feature vector and *sigmoid* the activation function. This defines a fully-connected layer for the network which joins the two Siamese twins and computes a distance metric over the feature space returning the similarity score between the two feature vectors.

The first Siamese twin returns the output feature vector for a given query molecule and the other returns an output feature vector for a molecule representing each one of the compound classes. This similarity measure is a probability, assuming a value between 0 and 1. If *score* is equal to 1, the probability of both compounds belonging to the same class is maximum. If *score* = 0, this probability is minimum according to the learnt similarity rule.

N-way One-Shot Learning Classification

The reduced amount of biological data available for training led us to adopt a new strategy to predict novel compounds using the proposed model. A one-shot classification strategy is applied to demonstrate the discriminative power of the learned features.

The Siamese network earlier described accepts pairs of compounds from a small training set D with a given number of N examples of encoded matrices of equal dimension and label l:

$$D = ((d_1, l_1),...,(d_N, l_N)) \tag{4}$$

The data for classification is organized in pairs, one example from a support set S and the other from a test set T. The support set consists of set of molecules representing each class selected at random whenever a one-shot learning task is performed. The support set has compounds representing each one of the compound categories and the test set has the test molecule of unknown class provided for classification.

In order to access the ability to make accurate predictions, a test instance d_j of unknown class is selected. Knowing that only one instance in our support set corresponds to that same class, the objective is to predict that class l belonging to D as the label l_i of an instance d_i.

Note that for every pair of input twins, our model generates a similarity score between 0 and 1 in one-shot. Therefore, to evaluate whether the model is really able to recognize similar molecules and distinguish dissimilar ones, we use an N-way one shot learning strategy (Figure 1). The test molecule is compared to N different ones and only one of those matches the original input. Thus, we get N different similarity scores $\{score_1,..., score_N\}$ denoting the similarity between the test molecules and those on the support set. This process is repeated across multiple trials, the model accuracy being determined as the percentage of correct predictions. Thus, in each trial, the pairs are organized for validation so that the first pair is a pair of instances of the same class, with the remaining pairs formed by compounds of different classes. If the pair of compounds of the same class (the first pair) gets the maximum similarity score, the model prediction is correct.

Over multiple trials, in each one-shot task, the Siamese network predicts which of the compounds present in the support set S most closely resembles the given test molecule in the test set T. The prediction *pred* corresponds to the pair (d_i,d_j) that returns the highest similarity score $score(d_i,d_j)$ in a one-shot trial with d_i the test molecule and d_j the support set molecule,

$$pred(d_i, S) = \arg\max(score(d_i, d_j)),\ d_j \in S \tag{5}$$

It is possible to verify that increasing N, more challenging it becomes to obtain a correct prediction and lower is the accuracy of the model. This is due to the fact that it is more difficult to obtain the maximum similarity score for the first pair due to the presence of a greater number of pairs in comparison at each trial.

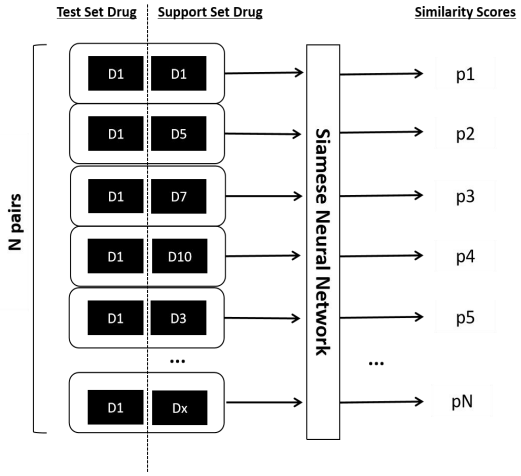

Figure 1: N-way one-shot classification using a Siamese neural network.

Results

In this paper, we propose a model able to predict novel compounds according to a degree of similarity between molecules. This metric is computed by a similarity function learnt by a one-shot Siamese neural network built upon two parallel and identical convolutional neural networks. To measure the model performance, the accuracy was determined using a N-way one-shot learning strategy described previously:

$$accuracy\ (\%) = \frac{number\ \ of\ \ correct\ \ prediction\ s}{number\ \ of\ \ trials\ \ per\ one-shot\ task} \tag{6}$$

The comparison of a given complex model with a set of simpler base models is a common strategy when assessing performance. Therefore, it was crucial to compare the proposed model with traditional machine learning approaches and simpler deep learning approaches (Table 1).

	N					
	2	3	4	5	7	10
Siamese Neural Network (validation)	94%	90%	84%	78%	70%	65%
Siamese Neural Network (training)	95%	92%	86%	84%	72%	70%
KNN	70%	55%	49%	43%	36%	30%
Naive Model	61%	43%	34%	31%	22%	19%
SVM	56%	42%	30%	24%	16%	12%
Random Forest	71%	58%	60%	44%	34%	20%
Multi-Layer Perceptron	76%	60%	36%	34%	22%	13%
Convolutional Neural Network	81%	70%	58%	46%	41%	39%

Table 1: N-way One-Shot Learning Accuracy Results.

References

[1] Altae-Tran, H., Ramsundar, B., Pappu, A. S., \& Pande, V. (2017). Low Data Drug Discovery with One-Shot Learning. ACS Central Science. https://doi.org/10.1021/acscentsci.6b00367.

[2] Lake, B. M., Salakhutdinov, R., Gross, J., \& Tenenbaum, J. B. (2011). One shot learning of simple visual concepts. In {Proceedings of the 33rd Annual Conference of the Cognitive Science Society}.

[3] Salakhutdinov, R. R., Tenenbaum, J. B., \& Torralba, A. (2012). One-Shot Learning with a Hierarchical Nonparametric Bayesian Model. JMLR Workshop and Conference Proceedings.

[4] A. Mayr, G. Klambauer, T. Unterthiner, and S. Hochreiter, "DeepTox: Toxic-ity prediction using deep learning,"Frontiers in Environmental Science, 2016.

[5] Koch, G. Siamese neural networks for one-shot image recognition. Ph.D. thesis, University of Toronto, 2015.

[6] N. R. C. Monteiro, B. Ribeiro, and J. Arrais, "Drug-Target Interaction Prediction: End-to-End Deep Learning Approach," IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2020.