

A comprehensive review on drug synergy prediction techniques.

Preeti karma*, Harmeet Singh

Department of Computer Science & Engineering, Sant Baba Bagh Singh University, Punjab, India

Abstract

Drug synergy is a critical area in the field of medicine. The correct drug-drug combination is highly needed to cure the life threatening diseases like blood cancer, lung cancer, throat cancer, etc. Drug synergy can be predicted with the help of machine learning models. In this paper, the comprehensive review on the recently proposed drug synergy techniques has been presented. The comparison between various machine learning based drug synergy prediction techniques has also been presented. The overall objective is to evaluate the various shortcomings in machine learning based drug synergy prediction techniques and to draw suitable future directions.

Keywords: Drug Synergy methods, machine learning, Prediction techniques of drug synergy.

Accepted on 05 October, 2020

Introduction

In the quest for clinical efficacy, drug combinations are a promising strategy in cancer treatment. Targeting a signaling pathway at one step may not be sufficient for reaching maximal effects on pathway inhibition. Using one agent at higher dose could be a short term solution. However, higher dose leads to increased toxicity and emergence of resistance to treatments. Resistance mechanisms to immunotherapy can occur by activation of compensatory signaling. For example, the activation of ERK signaling in melanoma when treated with BRAF inhibitors may lead to paradoxical activation of CRAF. Targeting BRAF and downstream

MEK at the same time proved to be beneficial for overall patient survival, by inhibiting the initial BRAF driver mutation and paradox CRAF activation. Alternatively to inhibiting two key proteins within the same pathway, a common strategy is to parallel inhibit two separate cancer pathways to maximize drug efficacy. For example, parallel inhibition of ERK and AKT could be beneficial as those pathways may be connected through cross talks and feedback loops in breast cancer. Given the enormous space of potential drug combinations, strategies to effectively predict their efficacy are highly desirable. Many methods predict drug synergy using chemical structure and genomic information. Preuer et al. used deep learning to predict synergy within the space of explored drugs and cell lines (Pearson's correlation of observed versus predicted synergy score $r=0.73$), but observed a much worse performance in predicting untested drugs ($r=0.48$) or untested cell lines ($r=0.57$). Jaeger et al identified new drug combinations using network topology of pathway cross-talk. However, gene mutation information, arguably the most actionable information in the clinic, was not used. In the recent Dialogue on Reverse-Engineering Assessment and Methods (DREAM) drug combination challenge, the best performing team used a protein-protein interaction to augment the genomic features based on their network distance from drug targets. Whilst the best performer achieved outstanding predictability comparable to the level of experimental replicates, synergy was predicted

based on supervised machine learning algorithms. A common bottleneck for the application of all supervised learning methods is the limited publicly available combinatorial drug screening data. In practice, the combinatorial explosion of drug pairs is the limiting factor to both the number of experimentally tested drugs, and the number of tested cell lines. Additionally, tested combinations are driven by expert's knowledge, and therefore may be focused on known biological examples and thereby bias the performance of supervised learning.

Drug synergy challenge

The similarities on the effect of drugs on gene expression were used to predict synergy. However, this requires the generation of expression data upon treatment, which is relatively costly. We here investigate if we can use similarities of single drugs in just the effect on cell survival to learn about the efficacy of combinations. We propose a methodology for prioritizing drug combinations and for cell line stratification based on the functional similarity between two target proteins. For this, we extend the notion of compound similarity to target similarity: the functional similarity of a pair of target proteins is defined as the correlation between the drug responses upon perturbation of those proteins, as a function of the activity of a set of essential pathways. Pathway activities are computed from data-derived gene sets, that have been demonstrated to be more predictive than pathway-based gene sets.

Different cancer types may be driven by different cancer pathways. Therefore, the similarity metric is context dependent. Two target proteins that are functionally very similar are likely to belong to the same signaling pathway; on the contrary, functionally dissimilar proteins are likely to belong to unrelated pathways. We find higher synergy likelihood when there is either very high or very low similarity. Based on this information, we build a compound prioritization methodology for high-throughput screens. Furthermore, we explore context specific (breast and colorectal cancer) drug combinations for their mode of actions based on known mechanistic insights

from the mono-therapies, to predict synergy and potentially enable patient stratifications in the clinic.

Conformal prediction

Traditional machine learning algorithms for classification problems simply predicts the class labels without any confidence. Conformal predictors expand on this as they output prediction regions for a confidence level provided by the user. The confidence value is an indication of how likely each prediction is of being correct, for example, a confidence of 95% implies that the percentage prediction error will be 5% on average. Conformal predictors are built on top of traditional machine learning algorithms, referred to as underlying algorithms, and they can be broadly categorized into transductive and inductive approaches; we refer to Papadopoulos (2008) for more details. We here consider the inductive approach called Inductive Conformal Prediction (ICP), which is more computationally efficient as compared with the transductive approach. In particular, we use Mondrian ICP with SVM or SVM+ as the underlying algorithms, and the SVM or SVM+ distance to the decision boundary to define the non-conformity measures (NCM). Mondrian conformal prediction has the advantage that we achieve validity for the individual classes. To evaluate the performance of conformal predictors, we consider the observed fuzziness, as defined in Vovk et al.

Synergistic drug pairs share distinct attributes

Drug synergy can arise due to a variety of diverse mechanisms which may present as distinct patterns in in vitro assays. Due to the large discrepancy between combination efficacy metrics we reasoned that each metric may be identifying different types of synergistic combinations. Therefore, we looked to quantify which drug attributes were shared among all synergistic drug pairs and which were metric specific. Since we have previously found drug structure to effect pharmacological attributes such as toxicity and molecular targets, we investigated if structure based similarity between drug pairs was indicative of a pair being synergistic. Using chemical fingerprint similarity, we found that synergistic pairs were more similar to each other than antagonistic and other non-synergistic drugs, across all metrics combination efficacy metric, drug synergy scores increased with drug structure similarity. These results run counter to expectations that synergistic drugs target distinct pathways. We further investigated drug attributes that could characterize synergistic drug pairs and focused on molecular targets of these drugs, which has also been shown to effect numerous other pharmacological attributes in past research. We found that drug pairs sharing a higher number of the same targets were not more synergistic or antagonistic than expected by chance.

Synergy metrics identify unique synergistic pathway combinations

Drug combination efficacy metrics principally vary in their intrinsic assumptions about drug synergy. Previous work in

cancer drug combinations has demonstrated that drug synergy can be achieved through a variety of pathway mechanisms ; therefore each metric is most likely identifying distinct pathway combinations. Using the KEGG database to identify pathways based on the molecular targets of each drug, we evaluated whether the combination of targeting two specific pathways was consistently synergistic or antagonistic for each metric. We specifically evaluated the pathway combinations which were most variable among metrics (i.e. were significantly enriched for synergy using some metrics and a loss of significance in other metrics). With the identification of these top differential pathway combinations we investigated the potential causes for the variability between metrics.

Suite of metric specific models predicts drug synergy

Since the synergistic drug combinations showed distinct characteristics when compared to antagonistic or other drug pairs, we reasoned that using a computational approach we could build a classification model to predict drug synergy or antagonism based on the similarity of various pharmacological and genomic attributes. Due to the diverse nature of each combination efficacy metric we chose to build a set of classification models, each fit with the synergistic/antagonism labels found using a specific metric, to create a model toolbox. Additionally, to account for the cell line specificity of drug synergy noted in past research and found within our own data we used a multi-task learning approach, which utilizes the strength of transfer learning while accounting for differences in synergy mechanisms between cell lines/cancer.

Related Work

Karen et al.(2015); Our work finds itself at the intersection of two domains: computational methods for prediction of side effects caused by DDI and neural networks for graph-structured data. As such, a review of related advances in each area will be presented here. As manually examining drug combinations and their possible side effects cannot be done exhaustively, computational methods were first developed to identify the drug pairs which create a response higher than the additive response they would cause if they did not interact.

Di Chen et al.(2016); This was previously done by framing the task as a binary classification problem and designing machine learning models (naïve Bayes, logistic regression, support vector machines) which predict the probability of a DDI, using the measurement of cell viability.

Mukesh et al.(2014); Other related approaches considered dose-effect curves or synergy and antagonism . An alternative way of approaching the task is provided by models which use the assumption that drugs with similar features are more likely to interact .

Assafet al.(2012); Using features such as the chemicals' structures, individual drug side effects and interaction profile fingerprints, the models use unsupervised or semi-supervised techniques (clustering, label propagation) in order to find DDIs. Alternatively, restricted Boltzmann machines and

However, all these methods are limited to either providing the likelihood of a DDI (but not its type if one exists), or lack applicability in inductive settings.

Bo Jin et al.(2017); the Multitask Dyadic Prediction in are two methods which overcome these challenges, and are thus going to be used as baselines against which our work will be compared. Multitask Dyadic Prediction is a proximal gradient method which uses substructure fingerprints to construct the drug feature representations. Similarly to our work, it has access. only to the chemical structure of the drug. Decagon, on the other hand, improves predictive power further by including additional relational information with protein targets of interest. Specifically, Decagon leverages this information by applying a graph convolution neural network architecture over a graph

Corresponding to the interactions between pairs of drugs, pairs of proteins and drug-protein pairs, treating discovery of novel DDIs as a link prediction task in the graph. While the protein-related auxiliary information is highly beneficial for the algorithm to use, it could also be expensive to obtain.

Joan et al.(2013); Compared to previous methods, our contribution is a model which learns a robust representation of drugs by leveraging joint information early on in the learning process. This allows it to bring an improvement in terms of predictive power, while maintaining an inductive setup where the model indicates the types of the possible

side effects by just looking at chemical structure of the drugs. Our model builds up on a large existing body of work in graph convolutional networks

Andreea et al.(2018) have substantially advanced the state-of-the-art in many tasks requiring graph-structured input processing (such as the chemical representation of the drugs leveraged here). Furthermore, we build up on work proposing co-attention as a mechanism to allow for individual set-structured datasets (such as nodes in multimodal graphs) to interact. Overall, these (and related) techniques correspond to one of the latest major challenges of machine learning with transformative potential across a wide spectrum of potential applications (not only limited to the biochemical domain).

Goodfellow et al.(2016); Many cancers have specific molecular causes, e.g., mutations in genes involved in the hallmark processes of cancer .Targeted cancer drugs directly affect those particular cancer genes. However,

the efficacy of a drug to block cancer cells' growth may be determined by additional genes. For example, trastuzumab is a human epidermal growth factor receptor-2 (HER2) antibody targeting HER2-overexpressed breast

cancer cells.

William et al.(2017); Thus it is valuable to model multiple cancer drugs sensitivity and data jointly.

In recent years, several groups and consortia have developed big datasets which include large-scale ex vivo pharmacological

profiling of cancer drugs and the genomic information of corresponding cell lines.

The drug sensitivity data for some groups of drugs are expected to be correlated, due to their common targets and

Santiago et al.(2012) similar pharma dynamic behaviors. To analyze such data, one straightforward method is to use (penalized) linear regression methods, for example lasso regressing each drug on all molecular features in a linear manner. Lasso could select a few relevant features with nonzero regression coefficient estimates from a large number of features. But it cannot address the heterogeneity of different molecular data sources.

Boulesteix et al. (2017) introduced integrative ℓ_1 -penalized regression with penalty factors (IPF-lasso) to shrink the effects of features from different data sources with varying ℓ_1 -penalties, to reflect their different relative contributions. While lasso or IPF-lasso can be extended to multivariate regression to jointly model multiple drugs sensitivity, the correlation of drugs is not reflected in the penalization of regression coefficients.

Diederik et al. (2014) proposed tree-lasso to estimate structured sparsity of multiple response variables assuming a hierarchical cluster structure in the response variables. Each cluster is likely to be influenced by some common features, for which the effects are similar between correlated responses. In this article, we propose the IPF-tree-lasso which borrows the strength of varying penalty parameters from IPF-lasso and the cluster structure in multivariate regression from tree-lasso. Thus, IPF-tree-lasso can capture the different relative contributions of multiple omics input data sources and the group structure of correlated drug response variables. Since some targeted

Integrative Predicting Drug Sensitivity

Cancer drugs might have similar mechanisms, for example the same target gene or signaling pathway, then these drugs are likely to have correlated sensitivities. IPF-tree-lasso is likely to select common relevant molecular features of these correlated drugs, and accordingly to shrink their coefficients with similar penalty parameters. Elastic net is also compared here, because it considers the grouping effect of correlated features and the 2-penalty can improve the prediction performance over lasso. Additionally, we also formulate the integrative elastic net with penalty factors (IPF-elastic-net) model to provide an extension of the elastic net with varying complexity parameters as well as varying parameters. However, IPF-tree-lasso and IPF-elastic-net have more complicated penalty terms which might require new optimization algorithms. We use augmented data matrix, so that the original smoothing proximal gradient descent method and cyclical coordinate descent algorithm for lasso can be employed directly. As elastic net and IPF-type methods have multiple penalty parameters to be optimized, the standard grid search is computationally not efficient. Frohlich and Zell proposed an interval-search algorithm, the efficient Parameter Selection via Global Optimization (EPSGO), which is more efficient. The rest of the paper is organized as follows. In

Section 2, the standard penalized regression methods and their extensions with structured penalties are introduced briefly. Section 3 describes the simulation scenario based on multivariate responses and different types of features, and then we present the simulation results and discussion. In Section 4, the Genomics of Drug Sensitivity in Cancer data are used to compare the prediction performance

Review

Methods of support vector machines (SVM)

Support vector machines (Vapnik and Vapnik, are one of the most successful methods for classification in machine learning. One of the key concepts of SVM is the use of separating hyper-planes to define decision boundaries, and the optimal decision hyper-plane is a plane in a multidimensional space that separates between data points of different classes and that also maximizes the margin, separating the two classes. SVM uses the kernel trick to generate a high dimensional nonlinear representation of the input data examples where it performs the separation with a continuous separation hyper-plane, such that the distances of misclassified data examples from the hyper-plane are minimized. In this study, we use a classification SVM for training our classification models with a Radial Basis Function

Conformal prediction Traditional machine learning algorithms for classification problems simply predicts the class labels without any confidence. Conformal predictors expand on this as they output prediction regions for a specific confidence level provided by the user. The confidence value is an indication of how likely each prediction is of being correct, for example, a confidence of 95% implies that the percentage prediction error will be 5% on average. Conformal predictors are built on top of traditional machine learning algorithms, referred to as underlying algorithms, and they can be broadly categorized into transductive and inductive approaches; we refer to Papadopoulos (2008) for more details. We here consider the inductive approach called Inductive Conformal Prediction (ICP), which is more computationally efficient as compared with the transductive approach. In particular, we use Mondrian ICP with SVM or SVM+ as the underlying algorithms, and the SVM or SVM+ distance to the decision boundary to define the non-conformity measures (NCM). Mondrian conformal prediction has the advantage that we achieve validity for the individual classes. To evaluate the performance of conformal predictors, we consider the observed fuzziness, as defined in Vovk et al. (2005).

Drug synergy cell line data

The drug pair synergy data was downloaded via NCI-ALMANAC17 and refined to include drug pairs with enough publically available data. In total 3647 unique drug pairs in 60 cell lines were analyzed. Using the raw data provided in NCI-ALMANAC, the R package SynergyFinder36 Version 1.6.1 was used to calculate the Bliss, ZIP, HSA and Loewe synergy score for each drug pair. We categorized drug pairs as

“synergistic” for each metric if their scores were within the top 366 5% and “antagonistic” if the scores were in the bottom 66.67%.

Pathway analysis

All known drug targets were collected from Drugbank and matched 369 to KEGG pathways via the KEGGREST37 R package using a custom R script. A fisher’s exact test was used to find the Odd’s Ratio for targeted pathway combination likely to be marked as synergistic for each metric. The most variable pathway combinations were found by identifying all combinations that had at least one synergy metric with an Odd’s Ratio lower confidence interval above 1.5 and an Odd’s Ratio higher confidence interval lower than 1.

Feature collection

Compound Features for the 3,647 drug pairs, multiple compound similarity features were collected. Additionally, using their known drug targets as listed in DrugBank, we collected drug target similarity features as well. The feature, source and metric used to measure similarity is listed in

Supplementary table

The measures of similarity included but were not limited to Pearson Correlation, Jaccard Index and Dice Similarity. In cases where there was insufficient or missing information, features were imputed by using the median value for that feature in drug pairs with complete information.

Network features

We curated a biological network that contains 399 protein-coding genes, 6,679 drugs, and 170 TFs. The protein-protein interactions represent established interaction, which include both physical (protein-protein) and non-physical (phosphorylation, metabolic, signaling, and regulatory) interactions. The drug-protein interactions were curated from several drug target databases.

Predictive model suite

Our predictive models were trained as binary classifiers using the features described above on the NCI ALMANAC data, with synergistic and antagonistic drug pairs being our respective classes. Every model included the same features, however the classes were determined by one of the five

Drug synergy measures (HSA, Bliss, Loewe, ZIP, ALMANAC Score). Multitask extremely randomized tree models, a decision tree model, was used after model selection and implemented using the R statistical software with the extra Trees package, the cancer cell line was used as each task. To evaluate predictive power 10-fold cross validation was used for each model. Down sampling was the chosen sub-sampling approach applied to each fold to account for the class imbalance between synergistic and antagonistic drug pairs.

Classification evaluation

For evaluating all the binary synergy classifications, receiver operating characteristic (ROC) and precision-recall curve (PRC) curves were created in R using the pROC41 and precrec packages respectively. Area-under-the-ROC curve (AUC) and area-under-the-PRC (AUPRC) scores were used to evaluate model performance.

DREAM Challenge validation data

Raw dose-response data from the DREAM-AZ Combination Prediction Challenge¹⁰ was used as an external dataset to test our models. We found 19 drug pairs, unseen by the models, available in the Challenge 1 data set tested within cell lines our models were trained on. For these 19 pairs features were collected in the same manner as described above and drug synergy scores for all metrics, besides the ALAMANAC score, were calculated as well. The correlation between all synergy scores were found using Pearson correlations. The synergy 415 scores were predicted using each model and then a Pearson correlation to the calculated scores were measured. Since the calculated HSA score was most significantly correlated with the given DREAM challenge score, the predicted HSA scores were used in the comparison to the DREAM challenge scores.

Shortcomings of existing techniques

After conducting the review on drug synergy prediction it has been found that the designing an efficient drug synergy prediction model is an ill-posed problem. Following gaps have been formulated after reviewing the existing techniques:

- 1 Ensemble learning: Majority of the existing researchers has utilized ensembling of different machine learning models to achieve higher accuracy rate. However, ensembling methods are computationally extensive in nature, and also, unable to achieve maximum accuracy.
- 2 Parameter tuning: Existing machine learning models suffer from the parameter tuning issues. It has been observed from the literature that the meta-heuristic techniques can be used to overcome the issue of parameter tuning in an efficient manner. However, majority of existing researchers have neglected the use of meta-heuristic techniques.
- 3 Pre-mature convergence: It has been observed that majority of existing meta-heuristic-based machine learning models such as particle swarm optimization, genetic algorithm etc. suffer from pre-mature convergence issue. It limits the performance of drug synergy prediction techniques.
- 4 Stuck in local optima: Majority of existing meta-heuristic-based drug synergy prediction models suffer from stuck in local optima issue.

- 5 Computational speed: The majority of existing meta-heuristic based machine learning models suffer from poor computational speed.

Conclusion

This paper presents the comprehensive review on drug synergy prediction techniques. After literature survey, it has been found that ensembling has been utilized to achieve higher accuracy rate. However, ensembling methods are computationally extensive in nature. Machine learning models may also suffer from the parameter tuning issues. Meta-heuristic techniques can be used to overcome the issue this issue. Some researchers have applied meta-heuristic techniques in the field of drug synergy prediction. However, these techniques suffer from pre-mature convergence, stuck in local optima, and poor computational speed. Therefore, the designing of an efficient drug synergy prediction model is still an open area of research.

References

1. Karen A. Ryall, Aik Choon Tan, Agrawal H. Systems biology approaches for advancing the discovery of effective drug combinations. *J Cheminform* 2015;1-7.
2. Di Chen, Huamin Zhang, Peng Lu, Xianli Liu, et al. Synergy evaluation by a pathway-pathway interaction network: a new way to predict drug combination. *Mol BioS*. 2016;12: 614–623.
3. Mukesh Bansal, Jichen Yang, Charles Karan, et al. A community computational challenge to predict the activity of pairs of compounds. *Nat biotech*. 2014;32:1213.
4. Assaf Gottlieb, Gideon Y Stein, Yoram Oron, et al. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol sys bio*. 2012;8:3-529.
5. Bo Jin, Haoyu Yang, Cao Xiao, et al. Multitask Dyadic Prediction and Its Application in Prediction of Adverse Drug- Drug Interaction. In *AAAI*. 2019;12:45-72.

*Correspondence to

Dr Preeti karma

Department of Computer Science & Engineering

Sant Baba Bagh Singh University

Punjab

India