Feature Compression for Predicting Effective Drug Combination

Guocai Chen[1], Xiaoqian Jiang[1], W. Jim Zheng[1]*

[1] School of Biomedical Informatics,
University of Texas Health Science Center, Houston, TX, USA

*Corresponding author: Wenjin.j.zheng@uth.tmc.edu

Abstract. Computational approach to predict effective drug combination can significantly improve drug efficacy while reducing drug toxicity. In this work, we employed a deep feature compression approach on gene expression data, pathway information and Ontology Fingerprints to improve the performance of a deep learning framework for effective drug combination. Our method indicates that the deep feature compression approach is an effective way to improve the performance of drug combination prediction.

Keywords: Feature compression, drug combination, Ontology Fingerprints

1 Introduction

Combined use of drugs may have extra synergistic effect that could lead to the reduced drug dosage and hence the toxicity. Computational methodologies to predict the effective drug combination make it possible to discover extensive drug combinations without expansive and time-consuming experiments. However, despite of our previous work to integrate ontology, literature and experimental data, the lack of experimental data for drug combination impeded the improvement of the accuracy of the computational methodologies for drug combination prediction, especially those methods based on deep learning approaches.

The performance of deep learning methods relies on the use of large amount of high quality, annotated data generated specifically for a specific task. However, the quantity of high quality training data for a targeted problem is often limited in most of the circumstances such as for drug combination prediction. This issue has been explored in many ways. Transfer learning [1] is one of the approaches to deal with the lack of training case problem where deep models are trained with relevant training data that has been well-studied and are largely available and is then amended and re-trained for the targeted task with small amount of data. The prerequisite of the use of Transfer learning is the availability of large amount training data in the relevant field. Another method to enable the effective use of deep learning approach for small training dataset is deep

feature compression [2-4]. By reducing the dimension of the inputs in the feature domain, feature compression decreases the number of weights that need to be trained for the deep model and thus the need of large amount of training data.

Several machine learning methods have been applied for the detection of effective drug combination in the AstraZeneca-Sanger Drug Combination Prediction DREAM Challenge (www.synapse.org/#!Synapse:syn4231880, DREAM2015) [5, 6], including regression, decision trees, random forests, Gaussian processes, SVM, neural networks, text mining, mechanistic network-based and others. Preuer K. et al [7] designed a feed-forward neural network with the integration of the heterogeneous resources as input to predict the drug synergy. Janizek J. introduced an extreme gradient boosted based approach, TreeCombo [8] to predict synergy of novel drug combinations.

Built upon our previous work employing a Stacked Restricted Boltzmann Machine to predict effective drug combinations from ontology, literature and experimental data [6], we applied deep feature compression on the input features for this deep belief network and significantly improved the performance of this model for the drug combination prediction.

2 Methods and Data materials

2.1 Data

The data used in this project include Ontology Fingerprints derived from Gene Ontology and literature, transcription profiling data and the drug sensitivity data, as described in our previous work [6]. The training datasets from experiments we used in this project contains 2199 pairs of drugs for 83 cell lines, which is sourced from DREAM2015 (www.synapse.org/#!Synapse:syn4231880). The features are compiled with Ontology Fingerprints [9-13], KEGG pathways [14-17] and gene expression data provided by DREAM2015.

2.2 Methods

Feature compilation

As described previously [6], features are compiled specifically for each cell line for effective genes. Each feature is a combination of the rank of the targeted gene for the cell line normalized with the minmax algorithm, the Ontology Fingerprint similarity [10-13, 18] and the inverse distance of the targeted genes in the extended KEGG pathway measured with InfoMap [19].

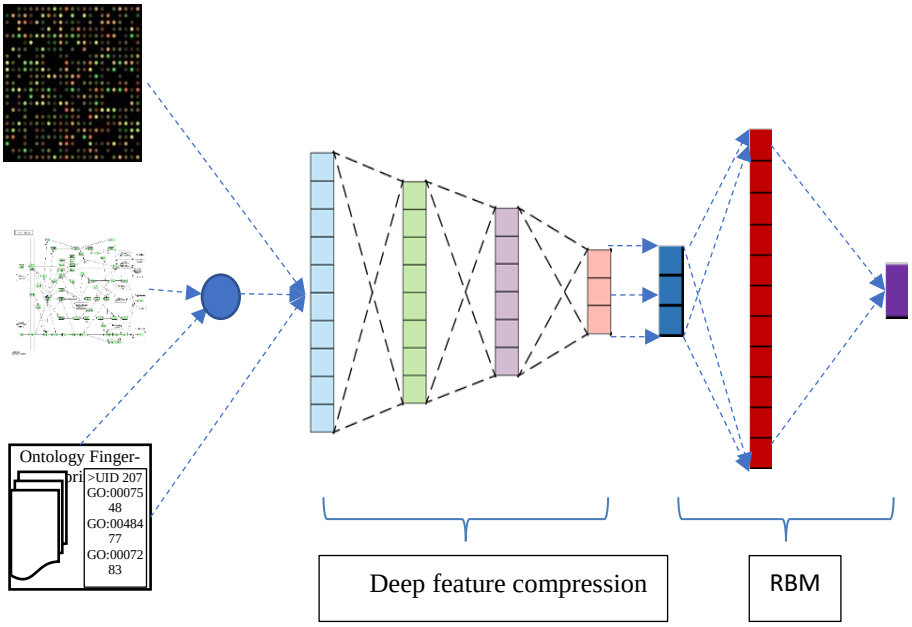

Fig. 1. A schematic diagram to show the workflow of employing deep feature compression to improve the performance of drug combination prediction. On the left is the three input data types: transcription profile (top), pathway (middle) and Ontology Fingerprints (bottom). In the middle is the deep feature compression layers, which will provide input to the Restricted Boltzmann Machine on the right to predict effective drug combination.

Feature compression and Deep learning with RBM

The vectors for the drug combination in the feature domain are fed into a deep Autoencoder as shown in Fig 1. The output in the middle layer is then extracted as the representative feature. These representative features are then used as the inputs to the Restricted Boltzmann Machine [20] for the drug combinations prediction.

Experiment and Evaluation

The model is trained and evaluated in 3 ways:

1. One model for one cell line.
2. One model for all cell lines with feature compressed independently for each cell line.
3. One model for all cell lines.

The datasets are evaluated using leave one out validation method.

A typical combination of hyperparameters we used is:

For deep Autoencoder, three hidden layers were used. The size of each layer is one 10^{th} of the size of the previous layer. If the size of the layer is less than 10, then double it. The other parameters are all default values.

For the Stacked RBM model we used 3 layer neural networks as well – the dimensions are the input size, 60 and 1 respectively. The other parameters are Weight cost 0.0001, Drop-out rate 0.5, Step ratio 0.01 Batch size 100. The SparseQ and SparseLambda may be slightly adjusted to balance the precision and recall.

3 Results

Comparing with the result we obtained previously [6]—precision 71.5%, recall 60.2%, f score 65.4%, as shown in table 1 the improved results are significantly better after applying deep feature compression. For method I, the overall precision is 77.1%, recall is 68.0% and f score is 72.2%. The f scores for 43 out of 83 cell lines are greater than 70% where we only had 32 out of 83 cell lines reported previously [6]. For method II, the overall precision is 73.3%, recall is 55.5% and f score is 63.2%. For method III, the overall precision is 72.7%, recall is 58.3% and f score is 64.7%. While the methods II and III show similar results as previously reported, we believe this is due to the sharing of a single model for all 83 cell lines. The diversity of these cell lines makes our approaches used in method II and III not very effective in prediction.

Table 1. Comparison of the performance

Methods	Precision	Recall	F-Score
Chen, Tsoi et al. 2018 [6]	71.5%	60.2%	65.4%
Best in Round 1*	34%	71%	46%
Best in Round 2*	38%	65%	48%
Best in Round 3*	42%	55%	48%
Method 1	**77.1%**	**68.0%**	**72.2%**
Method 2	73.3%	55.5%	63.2%
Method 3	72.7%	58.3%	64.7%

* The best performer in DREAM2015

4 Conclusions

We applied deep feature compression for the purpose of predicting effective drug combination. Our results indicate that reducing the dimension of the feature domain by deep feature compression can significantly improve the performance of the deep learning model we previously developed to predict effective drug combination [6].

5 References

1. Torrey, L. and J. Shavlik, Transfer learning, in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. 2010, IGI Global. p. 242-264.
2. Watanabe, S., Feature compression, in Advances in information systems science. 1970, Springer. p. 63-111.
3. Choi, J., et al. Context-aware deep feature compression for high-speed visual tracking. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
4. Choi, H. and I.V. Bajić. Deep feature compression for collaborative object detection. in 2018 25th IEEE International Conference on Image Processing (ICIP). 2018. IEEE.
5. Menden, M.P., et al., Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nature communications, 2019. **10**(1): p. 2674.
6. Chen, G., et al., Predict effective drug combination by deep belief network and Ontology Fingerprints. Journal of biomedical informatics, 2018.
7. Preuer, K., et al., DeepSynergy: Predicting anti-cancer drug synergy with Deep Learning. Bioinformatics, 2017. **1**: p. 9.
8. Janizek, J.D., S. Celik, and S.-I. Lee, Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. bioRxiv, 2018: p. 331769.
9. Chen, G., et al., Using Ontology Fingerprints to disambiguate gene name entities in the biomedical literature. Database, 2014. **bav034**: p. doi: 10.1093/database/bav034.
10. Qin, T., et al., Signaling network prediction by the Ontology Fingerprint enhanced Bayesian network. BMC Systems Biology, 2012. **6**(Suppl 3): p. S3.
11. Qin, T., et al., Finding pathway-modulating genes from a novel Ontology Fingerprint-derived gene network. Nucleic acids research, 2014. **42**(18): p. e138-e138.
12. Tsoi, L.C., et al., Evaluation of genome-wide association study results through development of ontology fingerprints. Bioinformatics, 2009. **25**(10): p. 1314-1320.
13. Tsoi, L.C., et al., Consistent Differential Expression Pattern (CDEP) on microarray to identify genes related to metastatic behavior. BMC Bioinformatics, 2011. **12**: p. 438.
14. Fabris, F. and A.A. Freitas, New KEGG pathway-based interpretable features for classifying ageing-related mouse proteins. Bioinformatics, 2016. **32**(19): p. 2988-2995.
15. Kanehisa, M. and S. Goto, KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 2000. **28**(1): p. 27-30.
16. Kanehisa, M., et al., KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic acids research, 2009. **38**(suppl_1): p. D355-D360.
17. Kanehisa, M., et al., The KEGG resource for deciphering the genome. Nucleic acids research, 2004. **32**(suppl_1): p. D277-D280.
18. Chen, G., et al., Using Ontology Fingerprints to disambiguate gene name entities in the biomedical literature. Database, 2015. **2015**: p. bav034.
19. Burk, C., F. Horton, and C. Amat, Infomap: a complete guide to discovering corporate information resources. Revista Española de Documentación Científica, 1994. **17**(2): p. 235.
20. Fischer, A. and C. Igel. An introduction to restricted Boltzmann machines. in Iberoamerican Congress on Pattern Recognition. 2012. Springer.
21.