Editorial

Making deep learning models transparent

Lujia Chen, Xinghua Lu

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA

Correspondence to: Lujia Chen. Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA. Email: xinghua@pitt.edu.

Comment on: Ma JZ, Yu MK, Fong S, *et al.* Using deep learning to model the hierarchical structure and function of a cell. Nat Methods 2018;15:290-8.

Received: 01 July 2018; Accepted: 14 July 2018; Published: 19 July 2018.

doi: 10.21037/jmai.2018.07.01

View this article at: http://dx.doi.org/10.21037/jmai.2018.07.01

Deep learning

Deep learning models (DLMs) have made groundbreaking advances in many artificial intelligence (AI) fields such as speech recognition, image analysis, and game playing (1,2). As biomedical research and medicine march into a big data era, it is foreseeable that DLMs will play an increasing important role in analyzing biomedical data.

Early-stage DLMs, often referred to as artificial neural network (ANN), were inspired by biological brain information processing. Although nowadays DLMs constitute a broader family of machine learning methods, the most common form of DLM is deep neural networks (DNNs), which contain one visible layer, multiple hidden layers and one output layer (if supervised). Each hidden layer consists of a set of hidden nodes ("neurons") fully connected to the nodes in adjacent layers, and the hierarchical hidden layers are believed to represent statistical structures of different degree of abstractions. Interestingly, as we have limited knowledge of how neurons in human brain acquire and store knowledge, we generally do not know what the so-called "neurons" in DLMs encode and how they learn a mapping function from inputs to outputs. In other words, contemporary DLMs largely behave as "black boxes".

The needs of visible DLMs

Impressed by the great performance of DLMs, more and more researchers focus on the latent variables involved in the neural networks and how they learn more accurate classifications (3). It's not only interesting to analyze the input-output relationship of the classifier, but also to look at what is going on inside the hidden layers of the network. The "transparent" DLMs could help us understand how the model processes signals, explain the predictions and gain the reason of task failure. This is especially true in the biomedical field when researchers not only seek for the model with the best prediction accuracy but also with the best biological explanation. For example, when applied to mining cancer omics data or systematic perturbation data, researchers may want to know the biological mechanisms that lead to cancer and use such knowledge to guide prescription of effective drugs.

Several visualization techniques have been used to understand and visualize DLMs in the field of image recognition such as layer activation, filter visualization and image occlusion (4). However, it's more challenging to "visualize" the hidden representation of biological data, because it is difficult for human to process patterns that do not have familiar visual cues. Therefore, there is an urgent need of visible neural networks (VNNs) that could make the hidden representations "transparent" from the perspective of biology to provide investigators a direct view of what hidden representations stand for (5). The VNNs will contribute to the cognition and development of DLMs suited for biomedical research. Designing VNNs for biomedical data could take advantage of the availability of prior knowledge, which could be encoded into the model to make the network apparent from the perspective of biology. For example, to model the cell signaling transduction system, the prior knowledge of the gene ontology (GO) could be used (6). Chen *et al.* showed that the deep belief network (DBN) could use hierarchical structure to capture the statistical patterns embedded in transcriptomic data corresponding to the biological signaling system. Furthermore, with the prior knowledge

of yeast transcription factor (TF)-gene interaction database and GO database, latent variables could be mapped to TFs and GO terms to make the network visible (7).

An example of VNN

More recently, Ma *et al.* developed a DLM, called DCell, to study how gene interactions impact cell growth (8). DCell is a VNN capable of inferring the biological representation between gene-disruption genotypes (single/pairs of gene deletion) and cell-growth phenotypes (cell growth) through a hierarchy of cell subsystems in the context of the ontology of cellular systems. The model mimics the biological processes of the deletion of a gene or pair of genes, propagated through the cellular hierarchy, to impact the parent subsystems containing them and finally lead to functional changes in small complexes, signaling pathways, organelles and ultimately a predicted cell growth phenotype. To visualize the "blackbox" between the input and the output, prior knowledge is incorporated into the model. The subsystems and their hierarchical relationship of DCell are based on Gene Ontology (GO) and the Clique-eXtracted Ontology (CliXO) database (9), which provide a hierarchy of biological concepts scaling from genes to proteins to organelles to whole cells, and including structures, functions, and hypotheses. DCell enables the authors to "visualize" latent variables by labeling them using corresponding GO and CliXO terms. The DCell captures system structure by inferring the biological information of the hidden representation and studying the mechanisms leading to the outcome of variations in phenotype.

The architecture of DCell could be thought of as a combination of supervised and unsupervised learning. It not only predicts cell growth (discrimination) but also infers the state of the subsystems (transparency) by simulation of DCell embedded in the structure. One advantage of the DCell is the sufficient amount of training cases compared with most biomedical tasks. DCell uses the big data of effects on yeast cell growth phenotype of deletion of each of 23 million pairs of genes (10), which effectively reduces the risk of over-fitting.

DCell could explain a genotype-phenotype association. To understand how deletion of a given pair of genes affects cell growth, Ma *et al.* examined the output of each subsystem when DCell input was set to related gene deletion, and quantified and prioritized which subsystems (GO terms) contributed most to DCell output phenotype using a designed metric called relative local improvement in predictive power (RLIPP). To make the DCell accessible for researchers to query genes of interest, Ma *et al.* developed an interactive website, http://d-cell.ucsd.edu, that visualizes the explanation of the growth phenotype (the activated cell subsystems) for any single or pair of yeast genes.

Same with every new model, it's important to validate and interpret the simulated results and prove its efficacy. Ma *et al.* used several examples to show that the predictions inferred from DCell are testable. For example, the simulated genotype (PMT1&IRE1)-phenotype (negative genetic interaction) association and genotype (REV7&RAD57)-phenotype (slow growth) association successfully agree with the findings from independent experiments. Ma *et al.* also showed that DCell had the ability of discovering new biological processes by validating previously undocumented CliXO terms.

In summary, DCell makes exciting progress in discovering the statistical pattern in gene interaction datasets and making the DLMs visible from the perspective of biological processes. It is a comprehensive model with the incorporation of DLMs, knowledge bases, and visualizations.

The future applications of DLMs in biomedical field

Nowadays, biomedical fields are collecting unprecedented amounts of data. Deep learning as a popular artificial-intelligence method provides a powerful tool for surveying and classifying biomedical data. In particular, the hierarchical organization of hidden nodes in DLMs closely reflect the hierarchical and compositional organization of cellular signaling systems, and their utilities remain to fully exploited. In the last decade, researchers have applied DLMs to biomedical fields, such as biomedical image analysis, genomics, proteomics, chemoinformatics, and drug discovery (11,12). Even though there are many advanced DLMs, such as CNN and generative adversarial network (GAN), most successful studies in biomedical fields selected architectures that suited to the problems at hand. For example, for a position-sensitive task such as medical image analysis, DLMs including CNN, ResNet and GoogleNet could be used to perform the classification task. For a task that is not position-sensitive, such as learning the statistical patterns of hierarchical cell signaling systems embedded in transcriptomic data, the unsupervised deep belief network (DBN) appears to be more appropriate.

Although the challenges remain, we foresee that with

more diverse systematic perturbation data from large projects, such as The Cancer Genome Atlas (TCGA) and the Library of Integrated Network-Based Cellular Signatures (LINCS) projects, we anticipate that novel DLMs and algorithms that fully take advantage of all available data to derive VNN will be developed, and such development will significantly advance biomedical research, with the potentials to transform research at both bench and bedside in several areas.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Medical Artificial Intelligence*. The article did not undergo external peer review.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jmai.2018.07.01). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

doi: 10.21037/jmai.2018.07.01
Cite this article as: Chen L, Lu X. Making deep learning models transparent. J Med Artif Intell 2018;1:5.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
2. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484.
3. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In European conference on computer vision. Springer, Cham 2014, pp 818-33.
4. Yosinski J, Clune J, Nguyen A, et al. Understanding neural networks through deep visualization. arXiv preprint arXiv:150606579 2015. Available online: https://arxiv.org/abs/1506.06579
5. Yu MK, Ma J, Fisher J, et al. Visible Machine Learning for Biomedicine. Cell 2018;173:1562-5.
6. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res 2015;43:D1049-56.
7. Chen LJ, Cai CH, Chen V, et al. Learning a hierarchical representation of the yeast transcriptomic machinery using an autoencoder model. BMC Bioinformatics 2016;17 Suppl 1:9.
8. Ma JZ, Yu MK, Fong S, et al. Using deep learning to model the hierarchical structure and function of a cell. Nat Methods 2018;15:290-8.
9. Kramer M, Dutkowski J, Yu M, et al. Inferring gene ontologies from pairwise similarity data. Bioinformatics 2014;30:34-42.
10. Costanzo M, VanderSluis B, Koch EN, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science 2016;353.
11. Mamoshina P, Vieira A, Putin E, et al. Applications of Deep Learning in Biomedicine. Mol Pharm 2016;13:1445-54.
12. Gawehn E, Hiss JA, Schneider G. Deep Learning in Drug Discovery. Mol Inform 2016;35:3-14.