Deep Learning Shines New Hopes on Solving the Half-a-Century-Old Problem of Protein Folding

Badri Adhikari
adhikarib @ umsl.edu
Assistant Professor of CS
Department of Mathematics & Computer Science
University of Missouri-St. Louis
Topics

- What is protein contact prediction?
- Early approaches (Feedforward NNs)
- Recent methods (ConvNets)
- How Deep Learning Contributes
- Conclusions
A Hypothetical Problem

- We would like to predict the meaning of a sentence ...
  - Using machine learning
A Hypothetical Problem

- We would like to predict the meaning of a sentence …
  - Using machine learning

- How to represent meanings?

Sentence

“Everyone should learn how to program because it teaches you how to think.”

Meaning

?
How to Represent the Meaning of a Sentence?

“Everyone should learn how to program because it teaches you how to think.”

Strong Connection
How to Represent the Meaning of a Sentence?

“Everyone should learn how to program because it teaches you how to think.”

Calculate Pairwise Connection Strengths
Machine Learning to Predict the Meaning of a Sentence

Inputs: Sentences (DB)

Neural Network

Outputs: Meaning (DB)

Calculate error
What is Protein Contact Prediction?

“Everyone should learn how to program because it teaches you how to think.”

Protein Sequence

Predict which amino acids interact with which...

Distance Map

English Sentence

G F G C N G P W D E D D M
What is Protein Contact Prediction?

“Everyone should learn how to program because it teaches you how to think.”

"G F G C N G P W D E D D D M"

Protein Sequence

Distance Map

Contact Map

Predict which amino acids interact with which..
Why Predict Contacts?

Precise protein contact prediction

Leads to..

Accurate protein structure / function prediction

Leads to..

Curing diseases through drug design
(cancer, mental health diseases)

Better understanding of how life works
(through understanding of how proteins work)

Improvements in Machine / Deep Learning
(because contact prediction is a hard problem)
Early Approaches to Contact Prediction
Feature Engineering

“Everyone should learn how to program because it teaches you how to think.”

<table>
<thead>
<tr>
<th>Character Length</th>
<th>8</th>
<th>6</th>
<th>5</th>
<th>3</th>
<th>2</th>
<th>7</th>
<th>7</th>
<th>2</th>
<th>7</th>
<th>3</th>
<th>3</th>
<th>2</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun (or not)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pronoun (or not)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| Polar (or not)   | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| Positively charged (or not) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hydrophobic (or not) | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| Helical (or not)  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

<table>
<thead>
<tr>
<th>G</th>
<th>F</th>
<th>G</th>
<th>C</th>
<th>N</th>
<th>G</th>
<th>P</th>
<th>W</th>
<th>D</th>
<th>E</th>
<th>D</th>
<th>D</th>
<th>D</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Approach: Consider Each Contact as a Separate Problem

How likely are these two close?

| Polar (or not) | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| Positively charged (or not) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hydrophobic (or not) | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| Helical (or not) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Pros:
- Many subproblems!
- Feasible to train
- More data to train

Predict Contact Map

Predict Contacting Pairs (instead of a full map)

Pair In Contact?

$G_1 - F_2 \ Y$

$G_1 - G_3 \ N$

$G_1 - C_4 \ N$

$G_2 - E_{10} \ Y$

$\ldots$

$\ldots$
Training and Testing

G

F

C

P

D      D

M
Training and Testing

How related are two words at position 3 and 10?
Early Approaches

The predictor neural network is a **standard feed-forward network, with 56 inputs** as given above, **ten hidden units**, and a **single output**.

**PROFcon**, a new method for predicting inter-residue contacts through a simple neural network. We considered **information from two ‘windows’ around two residues i and j** for which the probability of a spatial contact was predicted... We used 738 input, **100 hidden** and 2 output units (contact/non-contact).

We develop a new contact map predictor (SVMcon) that **uses support vector machines** to predict medium- and long-range contacts... SVMcon integrates profiles, secondary structure, relative solvent accessibility, contact potentials, and other useful features...
Standard Feed-forward Neural Networks Worked for $2^4$ Years

.. Classifiers are classic feed-forward neural networks, with 55 hidden units and a single output unit..

.. To train these very large networks, alternate rounds of offline and online training are carried out until no further improvement in accuracy is obtained..

Winner in CASP Competition (2012)

MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins

Winner in CASP Competition (2014)
Recent Methods
In 2016, Jinbo Xu’s Group Tested ConvNets...

.. By stacking **multiple convolution layers**, the network can **learn information in a very large sequential context**..

.. test results suggest that **deep learning can revolutionize protein contact prediction**...

![Feature visualization](image)

**Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model**

Sheng Wang, Shijun Sun, Zhen Li, Renyu Zhang, Jinbo Xu

doi: https://doi.org/10.1016/j.bios.2019.02.003

*Now published in PLoS Computational Biology.* doi: 10.1371/journal.pcbi.1005324

**Winner in CASP Competition**
*(2016 & 2018)*
A single ConvNet is Much More Accurate than Feed-forward NN

..Using pair frequency data as the input features, DeepCov is able to almost match the average performance of MetaPSICOV2, which is quite impressive given the simplicity of the feature set.

..With the same features as input, a CNN network trained with all contacts and non-contacts achieves a slightly better precision of 35.4% on top L/5 long-range contacts than DNCON 1.0. So, a single CNN model performs better than a boosted and ensembled deep belief networks, suggesting that the deep convolutional neural network (CNN) is more suitable for contact prediction than the deep belief network (DBN).
How & Why Do ConvNets Work?

Artificial neurons are inspirations of biological neurons. 

**Convolutional neurons** are like our “eyes with memory”.

Eye’s understanding of the input data

1st Layer of Conv. Neurons
How & Why Do ConvNets Work?

Artificial neurons are inspirations of biological neurons.

**Convolutional neurons** are like our "eyes with memory".

1st Layer of Conv. Neurons

2nd Layer of Conv. Neurons
What ConvNet Architectures are Best Fit for This Problem?

Top Methods in the most recent CASP Competition

DeepMetaPSICOV (DMP) in CASP13
Shaun M Kandathil
University College London
&
The Francis Crick Institute

ResTriplet/TripletRes:
Learning contact-maps from a triplet of coevolutionary matrices
Eric W. Bell, Yang Li, Chengxin Zhang,
Dong-Jun Yu, Yang Zhang
Department of Computational Medicine and Bioinformatics,
University of Michigan - Ann Arbor

All these results show that residual networks are best architectures (for this problem)
What are Residual Networks?

Input Features → 1st Layer’s Feature Mappings → 2nd Layer’s FM → 3rd Layer’s FM

Information continues to be lost...
.. learning becomes difficult
What are Residual Networks?
What Variations of Residual Architectures are Best Fit?

- To obtain an answer we have to try ‘almost’ all possible architectures
  - A lot of computing resources (GPUs)

- The input data for training is [2 GB to 200 GB+]
  - In one epoch (less than 20 minutes) we need to read 200 GB of data
  - On HPC clusters such as Lewis, training takes at least 10 days (with regular hard-drives)
    - Great GPUs (V100) but poor time limits (2 hours) & slow HDDs
    - We need SSDs (SATA & M.2)

- Applied to Google for resources
  - $5000 worth of Google Cloud Credits
  - Finished them in less than a week and requested more

- Applied to NVIDIA for resources
  - Awarded a Quadro P6000 GPU (performs similar to V100s; extremely useful)
We Tested Various Residual Networks Architectures

(a) Residual Block

(b) Residual with Dropout

(c) Dilated Residual

(d) Dilated with Dropout

Input (256 x 256) x N channels

Output

+ [X_f]

Input (256 x 256) x N channels

Output

+ [X_{f+1}]

Input (256 x 256) x N channels

Output

+ [X_{f+1}]

Input (256 x 256) x N channels

Output

+ [X_{f+1}]

Input (256 x 256) x N channels

Output

+ [X_{f+1}]

Batch Normalization
ReLU
Conv 3x3 (64 filters)
Batch Normalization
ReLU
Conv 3x3 (64 filters)
Dropout
ReLU
Conv 3x3 (56 filters)
Batch Normalization
ReLU
Conv 3x3 (64 filters)
Batch Normalization
ReLU
Conv 3x3 (64 filters)
Dropout
ReLU
Conv 3x3 (56 filters)
Batch Normalization
ReLU
Conv 3x3 (64 filters)
Batch Normalization
ReLU
Conv 3x3 (56 filters)
Conv 3x3 Dilation = 1/2/4
Conv 3x3 Dilation = 1/2/4
Residual Networks with Dilation & Dropout Perform Best

..Here, we experiment with two diverse datasets that use different input features. When trained on the DeepCov dataset consisting of 3,456 proteins, using the same dataset for training and testing our method achieves up to 6% and 15% higher precision on the PSICOV150 protein dataset when top L/5 and L/2 long-range contacts are evaluated, respectively (L is protein length).
But.. Is There Room for Improvement? YES

- At the most recent CASP conference:
  “It was good to see Google DeepMind win this time..
  I was sick of seeing Rosetta win since almost two decades.”
  - a senior scientist at the conference

- Google plans to continue its ‘fundamental’ research

- We are still far from end-to-end deep learning
How Does Deep Learning Deliver Improved Performance?
Can We Learn to Predict Contacts WITHOUT ‘True’ Contacts?

MSEIITFPQQTVVYPEINVKTLSQAVKNIWRISHQQKSGIEIIQEKTLRISLYSRDLDEA

---NTLSQKENMYPEINIKAMNQAVNITIWLLAQRQTSGIEIINDKVKRISLYSREFDE-
-------------LTPPDTEVMKTARQNILHVTLALKLDFLPVMKEKMRPLQDALISADK-
-------------ILTPPDNEIMNDARQNILQASALKLDLDFLPVMKEKMLPLQTALKRADKV
MVVRNSAKIAIAEHSDDMAQINLKLLEGNRLLLETLQEQIDSITLRSAALESTMGEITA----
AGIARLGKLLDKVSALTYLDLIEERDRRLRLRYNALLEESRTAHQEEKATAAKLDELT--
-------------LTPPDNEVMETARQNILQVTALKLDLDFLPVMKEKMLPLQAALMSADKV
DVVRRSAQAIAAHSDDVAQINLTLLESNRLLETLQQQIDSMTQRSATQESAMALIAA----
AGIARLGKLLDKVSALNYLDLIEERDRRLRYNALLEDSRTALREAQATREKLDELTA

Does this mean we can write an algorithm to predict contacts?
Can We Learn to Predict Contacts WITHOUT ‘True’ Contacts?

Physical contacts

Observed correlations

Predicted contacts

Causative

Transitive

Protein structure prediction from sequence variation

Debora S Marks, Thomas A Hopf & Chris Sander

Can We Write Algorithms to Remove Transitive Noise?

Protein 3D Structure Computed from Evolutionary Sequence Variation
Debora S. Marks, Lucy J. Colwell, Robert Sheridan, Thomas A. Hopf, Andrea Pagnani, Riccardo Zecchina, Chris Sander
Published: December 7, 2011 • https://doi.org/10.1371/journal.pone.0028766

FreeContact: fast and free software for protein contact prediction from residue co-evolution
László Kaján, Thomas A Hopf, Matúš Kalaš, Debora S Marks and Burkhard Rost
BMC Bioinformatics 2014 15:85
Received: 30 September 2013 | Accepted: 18 March 2014 | Published: 26 March 2014

PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments
David T. Jones, Daniel A. Buchan, Domenico Cozzetto, Massimiliano Pontil
Author Notes
Bioinformatics, Volume 28, Issue 2, 15 January 2012, Pages 184–190,
https://doi.org/10.1093/bioinformatics/btr638
Published: 17 November 2011 Article history

CCMpred—fast and precise prediction of protein residue–residue contacts from correlated mutations
Stefan Seemayer, Markus Gruber, Johannes Söding
Author Notes
Bioinformatics, Volume 30, Issue 21, 1 November 2014, Pages 3128–3130,
https://doi.org/10.1093/bioinformatics/btu500
Published: 26 July 2014 Article history

2019 prepare.ai Conference April 9, 2019 Deep Learning & Protein Contact Prediction - Slide 33
Can Deep Learning Remove Transitive Noise?

**Precision of top L/5 Long-Range Contacts**

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEEPCON</td>
<td>60.0</td>
</tr>
<tr>
<td>DeepCov</td>
<td>50.0</td>
</tr>
<tr>
<td>PconsC4</td>
<td>40.0</td>
</tr>
<tr>
<td>CCMpred</td>
<td>30.0</td>
</tr>
<tr>
<td>FreeContact</td>
<td>20.0</td>
</tr>
</tbody>
</table>

**DEEPCON: Protein Contact Prediction using Dilated Convolutional Neural Networks with Dropout**

- Badri Adhikari
- doi: https://doi.org/10.1101/590455

This article is a preprint and has not been peer-reviewed [what does this mean?].
Conclusions

1) Groups who were good at exploring ‘new flavors’ did well
   - Learn various deep learning methods, even when you don’t see a direct fit to your problem

2) Balanced efforts of ML experts and domain experts brought success
   - Do you have enough ML ‘breadth’ or team?

3) When end-to-end is not possible, correct feature engineering becomes important
   - Is feature engineering solved for your problem? If not, focus your research here!
   - For example, for standard images, we don’t need feature engineering

4) Using ‘a lot of data’ and ‘deep architectures’ improves performance
   - Have you tried using “all the data” and “a large architecture”? 
Acknowledgements

Research Support & Contribution

Cezary Janikow  Sharlee Climer  Cynthia Jobe  Anthony Ackah-Nyanzu  Patrick Kong  Sri Harsha Akurathi

IT Support

Philip Reiss  Kenneth Voss  Michael Remier  MU - Research Computing Support Services (RCSS)

Computing Resources

Google  NVIDIA  University of Missouri System
THANK YOU