UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

Applying Saliency Map Analysis to CNNs on Protein Secondary Structure Prediction

by Guillermo Romero Moreno

The relatively new field of deep learning is slowly being transferred to the biology field and pushing its state-of-the-art achievements. However, improvements in performance come with the drawback of opaqueness, as what deep learning machines learn cannot be fully understood. Although a few authors have already started applying deep network interpretability techniques on biological problems to overcome this issue, none of them has been applied yet to the problem of protein secondary-structure prediction.

The aim of this work is to develop interpretability techniques for state-of-the-art deep networks that have been trained to solve the secondary-structure prediction problem. For doing so, a way to apply and aggregate saliency maps has been construed and applied to a near-state-of-the-art convolutional network, showing some further insights of the relationship between the inputs and the outputs. These results could be of double value: on one side, it may help biologists to get a better understanding on the underlying structural protein processes; on the other, machine learning researchers can understand better their machines and spot their flaws more easily.

Chapter 1

Results & Discussion

I have trained the network described in section **??** for 400 epochs with the same parameters as the ones used by Jurtz et al. (2017); i.e., gradient clipping at 20, regularization term $\lambda = 10^{-3}$, and training-validation split at the 5278th sequence. The resulting network reaches an accuracy of 67.7% on the test set, which is not to far from the 70% of the state-of-the-art (see section **??**). I believe that such a small difference does not undermine the effectiveness of the other results, while notoriously alighting the computational burden of a more complicated model.

1.1 Outlier analysis

In order to analyse the performance space a bit better, the average accuracy per sequence has been calculated and it has been plotted in Figure 1.1 with respect to the sequence length. The distribution exhibits the typical funnel shape that one could expect from processes with random variables forming groups of different sizes: the bigger the groups, the smaller the variance. The funnel ceases to shrink at length about 400, so it would be particularly interesting to understand why the network is classifying worse (60% and below) some of the sequences above that length.

If we observe the color scheme of the figure, we can understand right away that sequences rich in α-helix are generally better predicted than β-sheets and coils. An explanation could be that while α-helix sizes are up to CHECK NUMBER, which is inside the window size, β-sheets interact with amino-aids further away in the sequence, which is not possible to be captured with the window of the network, of lateral size of 9.

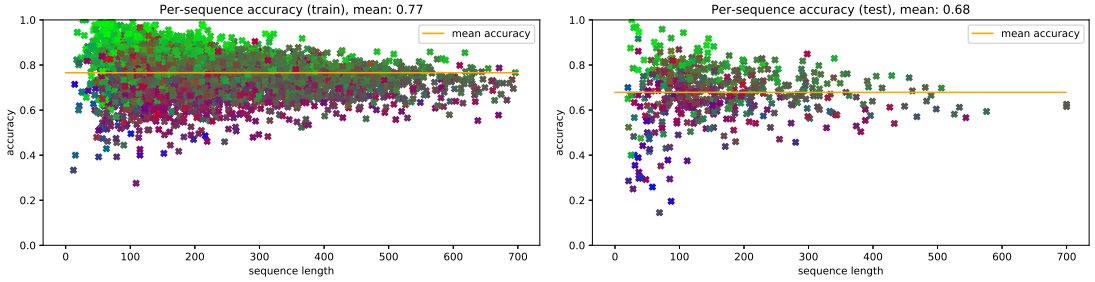

FIGURE 1.1: The mean accuracy per sequence by sequence length. The 5504 sequences of the training set are shown on the left and the 514 of the test set on the right. Each point represents a single protein, and its colour corresponds to the amount of β-sheets (red), α-helices (green), and coils (blue) it has. A purple point, for instance, would predominantly have β-sheets and coils.

1.2 Feature visualization

1.2.1 First layer filters

1.2.2 Saliency maps on layers?

1.3 Saliency maps on inputs

Before anything else, it is worth commenting that the saliency map outputs have many dimensions, since each position at each sequence has a saliency map with shape $8x42x19$, corresponding to the 8 classes (outputs), the 42 inputs and the total window size. The results can be shown in multiple different ways depending on which dimensions are preserved and which are aggregated.

Analysis on amino-acids and *pssm*

When looking at typical secondary-structure prediction algorithm, there is one point that may raise some suspicion: the inclusion of half of the inputs as one-hot encoded (amino-acids) and the other half as dense vectors (*pssm*). One could thing that this discrepancy may strongly favour the information coming from the dense part, since the weights associated to it will learn much faster in a typical gradient descent learning schema.

Saliency maps can be used to prove whether this hypothesis is right by inspecting which of the input groups is being most decisive in the classification process. For doing so, each saliency map is divided into two groups of $8x21x19$, and all the values inside added up to a single saliency score. Thus, each position of each sequence will have two scores, one for the amino-acids and one for the *pssm*.

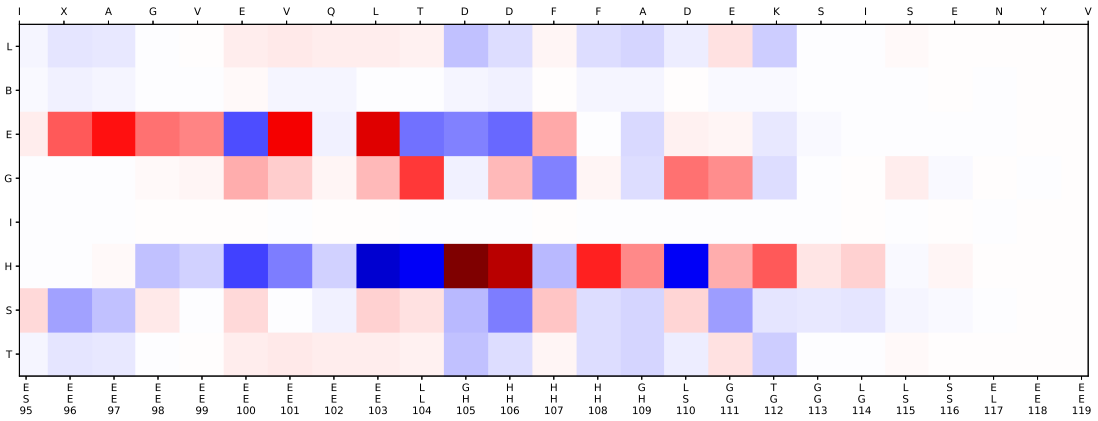

FIGURE 1.2:

Options (all options could be applied either to aa or to pssm, 6 dimensions left)

1.3.1 Saliency maps on single sample sequences

analyse per-class single sequence (sample-based) (4 dimensions left): aa-aggregated (3 dimensions) and added in the sequence position, either respecting each position (3 dimensions) or aggregating them in a heat map (2 dimensions), or aggregating the positions in a heat-map but containing the 8 classes (3 dimensions) (Colour the labels with the RGB either form Figure 1.1 or with the colour code from 1.3.2).

Have the x labels including amino-acid, prediction, and real label. Colour classes with the codes mentioned in the previous paragraph. Consider marking mismatches in bold.

Have as the sample one of the sequences with low accuracy and high length from 1.1

1.3.2 Sheer addition

aggregate all individual saliency maps sheer addition (4 dimensions left): class-aggregated (3 dimensions), aa-aggregated (3 dimensions), or class+aa-aggregated (2 dimensions)

1.3.2.1 Per-aminoacid and class aggregations

1.3.2.2 Per-class aggregations

First thing to notice: pssm is way more relevant than one-hot encoded aas. No wonder, it learns faster.

1.3.2.3 Per-aminoacid aggregations

1.3.3 Clustering techniques

aggregate all individual saliency maps clustering (5 dimensions left) Using the per-class window-aggregated version of individual saliency maps (4 dimensions left) Cosine distance metric. Show either all profiles per-cluster (3 dimensions), or aggregated profiles (2 dimensions) Show t-SNE with points coloured by cluster

Bibliography

Zikrija Avdagic, Elvir Purisevic, Samir Omanovic, and Zlatan Coralic. Artificial Intelligence in Prediction of Secondary Protein Structure Using CB513 Database. *Summit on translational bioinformatics*, 2009:1–5, 2009. ISSN 2153-6430.

Akosua Busia and Navdeep Jaitly. Next-Step Conditioned Deep Convolutional Neural Networks Improve Protein Secondary Structure Prediction. *arXiv:1702.03865v1*, 2017.

Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Wei Xie, Gail L. Rosen, Benjamin J. Lengerich, Johnny Israeli, Jack Lanchantin, Stephen Woloszynek, Anne E. Carpenter, Avanti Shrikumar, Jinbo Xu, Evan M. Cofer, David J. Harris, Dave DeCaprio, Yanjun Qi, Anshul Kundaje, Yifan Peng, Laura K. Wiley, Marwin H. S. Segler, Anthony Gitter, and Casey S. Greene. *Opportunities And Obstacles For Deep Learning In Biology And Medicine*. 2017. ISBN 0000000305396.

Dumitru Erhan, Yoshua Bengio, Aaron Courville, and Pascal Vincent. Visualizing higher-layer features of a deep network. *Bernoulli*, 2009.

C. Fang, Y. Shang, and D. Xu. A new deep neighbor residual network for protein secondary structure prediction. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 66–71, Nov 2017a.

Chao Fang, Yi Shang, and Dong Xu. MUFold-SS: Protein Secondary Structure Prediction Using Deep Inception-Inside-Inception Networks. sep 2017b.

Alejandro Fontal. *Neural Networks for Subcellular Localization Prediction*. PhD thesis, Wageningen University & Research, 2017.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. *PMLR*, 9:249–256, 2010. ISSN 15324435.

Leandro Takeshi Hattori, Cesar Manuel Vargas Benitez, and Heitor Silverio Lopes. A deep bidirectional long short-term memory approach applied to the protein secondary structure prediction problem. In *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6, 2017. ISBN 978-1-5386-3734-0.

Rhys Heffernan, Yuedong Yang, Kuldip Paliwal, and Yaoqi Zhou. Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. *Bioinformatics*, 33(18):2842–2849, 2017. ISSN 14602059.

Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. feb 2015.

Alexander Rosenberg Johansen, Casper Kaae Sønderby, Søren Kaae Sønderby, and Ole Winther. Deep Recurrent Conditional Random Field Network for Protein Secondary Prediction. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics - ACM-BCB '17*, 2017. ISBN 9781450347228.

Vanessa Isabell Jurtz, Alexander Rosenberg Johansen, Morten Nielsen, Jose Juan Almagro Armenteros, Henrik Nielsen, Casper Kaae Sønderby, Ole Winther, and Søren Kaae Sønderby. An introduction to deep learning on biological sequence data: Examples and solutions. *Bioinformatics*, 33(22):3685–3690, 2017. ISSN 14602059.

Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogenbonded and geometrical features. *Biopolymers*, 22 (12):2577–2637, 1983. ISSN 10970282.

Jack Lanchantin, Ritambhara Singh, Beilun Wang, and Yanjun Qi. Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks. 2016. ISSN 23356936.

Zhen Li and Yizhou Yu. Protein Secondary Structure Prediction Using Cascaded Convolutional and Recurrent Neural Networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016. ISBN 978-1-57735-770-4.

Zeming Lin, Jack Lanchantin, and Yanjun Qi. MUST-CNN: A Multilayer Shift-and-Stitch Deep Convolutional Architecture for Sequence-based Protein Structure Prediction. may 2016.

Christophe N. Magnan and Pierre Baldi. SSpro/ACCpro 5: Almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics*, 2014. ISSN 14602059.

Aravindh Mahendran and Andrea Vedaldi. Understanding deep image representations by inverting them. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015. ISBN 9781467369640.

Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, and Klaus Robert Müller. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition*, 2017. ISSN 00313203.

Grégoire Montavon, Wojciech Samek, and Klaus Robert Müller. Methods for interpreting and understanding deep neural networks, 2018. ISSN 10512004.

Alexander Mordvintsev, Michael Tyka, and Christopher Olah. Inceptionism: Going deeper into neural networks, google research blog, 2015.

Chris Olah, Alexander Mordvintsev, and Ludwig Schubert. Feature Visualization. *Distill*, 2017. ISSN 2476-0757.

Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning Important Features Through Propagating Activation Differences. apr 2017.

Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv.org*, 2014.

Søren Kaae Sønderby and Ole Winther. Protein Secondary Structure Prediction with Long Short Term Memory Networks. *arXiv:1412.7828*, 2014.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic Attribution for Deep Networks. 2017. ISSN 1938-7228.

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. dec 2013.

Sheng Wang, Jian Peng, Jianzhu Ma, and Jinbo Xu. Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. *Scientific Reports*, 6, 2016. ISSN 20452322.

Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014. ISBN 9783319105895.

Jian Zhou and Olga G. Troyanskaya. Deep Supervised and Convolutional Generative Stochastic Network for Protein Secondary Structure Prediction. 2014.

Jian Zhou and Olga G. Troyanskaya. Predicting effects of noncoding variants with deep learning-based sequence model. *Nature Methods*, 12(10):931–934, 2015. ISSN 15487105.

Jiyun Zhou, Hongpeng Wang, Zhishan Zhao, Ruifeng Xu, and Qin Lu. CNNH_PSS: Protein 8-class secondary structure prediction by convolutional neural network with highway. *BMC Bioinformatics*, 2018. ISSN 14712105.