Convolutional LSTM Networks for Subcellular Localization of Proteins

Machine learning is widely used to analyze biological sequence data. Non-sequential models such as SVMs or feed-forward neural networks are often used although they have no natural way of handling sequences of varying length. Recurrent neural networks such as the long short term memory (LSTM) model on the other hand are designed to handle sequences. In this study we demonstrate that LSTM networks predict the subcellular location of proteins given only the protein sequence with high accuracy (0.902) outperforming current state of the art algorithms. We further improve the performance by introducing convolutional filters and experiment with an attention mechanism which lets the LSTM focus on specific parts of the protein. Lastly we introduce new visualizations of both the convolutional filters and the attention mechanisms and show how they can be used to extract biologically relevant knowledge from the LSTM networks.

General information
State: Published
Organisations: Department of Systems Biology, Center for Biological Sequence Analysis, Functional Human Variation, Department of Applied Mathematics and Computer Science, Cognitive Systems, University of Copenhagen
Contributors: Sønderby, S. K., Sønderby, C. K., Nielsen, H., Winther, O.
Number of pages: 13
Pages: 68-80
Publication date: 2015

Host publication information
Title of host publication: Algorithms for Computational Biology: Second International Conference, AlCoB 2015, Mexico City, Mexico, August 4-5, 2015, Proceedings
Volume: 9199
Publisher: Springer
Editors: Dediu, A., Hernández-Quiroz, F., Martin-Vide, C., Rosenblueth, D. A.
ISBN (Print): 978-3-319-21232-6
ISBN (Electronic): 978-3-319-21233-3
Keywords: Subcellular location, Machine learning, LSTM, RNN, Neural networks, Deep learning, Convolutional networks
Electronic versions:
Convolutional_LSTM.pdf
DOIs:
10.1007/978-3-319-21233-3_6
Research output: Research - peer-review › Article in proceedings – Annual report year: 2015