Convolutional LSTM Networks for Subcellular Localization of Proteins

Nielsen, Henrik; Sønderby, Søren Kaae; Sønderby, Casper Kaae; Winther, Ole

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Nielsen, H., Sønderby, S. K., Sønderby, C. K., & Winther, O. (2015). Convolutional LSTM Networks for Subcellular Localization of Proteins. Abstract from First Annual Danish Bioinformatics Conference, Odense, Denmark.

23 Convolutional LSTM Networks for Subcellular Localization of Proteins.

Henrik Nielsen, Søren Kaae Sønderby, Casper Kaae Sønderby, Ole Winther.

Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark; The Bioinformatics Centre, University of Copenhagen, Denmark.

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.

24 The Bioinformatics Centre: Ongoing research and hiring opportunities.

Henriette Husum Bak-Jensen.

The Bioinformatics Centre, Section for Computational and RNA Biology, University of Copenhagen, Denmark.

25 Rhea, a manually curated resource of biochemical reactions.

Kristian B. Axelsen, Anne Morgat, Thierry Lombardot, Lucila Aimo, Anne Niknejad, Nevila Hyka-Nouspikel, Elisabeth Coudert, Nicole Redaschi, Lydie Bougeleret, Christoph Steinbeck, Ioannis Xenarios, Alan Bridge.

SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; Institute for Plant and Environmental Science, University of Copenhagen, Denmark.

Rhea (http://www.ebi.ac.uk/rhea) is a comprehensive and non-redundant resource of expert-curated biochemical reactions described using species from the ChEBI (Chemical Entities of Biological Interest) ontology of small molecules. Rhea is designed for the functional annotation of enzymes and the description of genome-scale metabolic networks, providing stoichiometrically balanced enzyme-catalyzed reactions, transport reactions and spontaneously occurring reactions. Rhea reactions are used as a reference for the reconciliation of genome-scale metabolic networks in the MetaNetX resource (www.metanetx.org) and serve as the basis for the computational generation of a library of theoretically feasible lipid structures in SwissLipids (www.swisslipids.org). Recent developments include the provision of reactions involving complex macromolecules such as proteins, nucleic acids and other polymers and substantial growth of Rhea through sustained literature curation efforts.